

Email: cgapbs-r@mail.nih.gov  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:1393813  
 High quality sequence stop: 439.  
 Location/Qualifiers

## FEATURES

source

1..587  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3513237"  
 /clone\_lib="Soares\_NMAX\_maxillary\_process"  
 /tissue\_type="maxillary process"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker: Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTGAAGTGGAGGCGCGCCGCTTTTTTTTTTTTTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 138 a 160 c 121 g 168 t

ORIGIN

## alignment\_scores:

Quality: 553.00 Length: 117  
 Ratio: 4.982 Gaps: 1  
 Percent Similarity: 94.872 Percent Identity: 89.744

## alignment\_block:

US-09-327-750d-33 x BG228077/rev ..

Align seg 1/1 to reverse of: BG228077 from: 1 to: 587

13 MetGluAsnAspHisGlnLysGluGluLysGluGluLysProGlnAs 29

587 ATGAAGATGACCATCAGAAAAAGCAGGAGGAAGAAATACCCACAAGA 538

29 pThrIleLysArgGluPro.ValValAlaProThrPheGluAlaGlyLys 45

537 TACCATCAGAAGGAGGAGCGCTGTGGCCCTGACCTCCGAGGCTGCACA 488

46 AsnCysAlaProArgGlyClyArgArgPheArgValArgGlnProI 62

487 AACTGTGCACCTAGAGAGGTCGACGCGGTTCCGGGTTCCGACGCCAT 438

62 eSerHisTyrArgTrpAspLeuMetHisArgValGlyGluProGlnGlyA 79

437 CGCTCACTATAGATGGGACCTGATCAGAGGGTTGGGAGCCCGAGGAA 388

79 rgMetArgGluAsnValGlnArgPheGlyGluAspMetArgGlnLeu 95

387 GGATGAGAGAGGAGACCTACAGAGGTTGGGGGTGATGTGAGACAGCTC 338

96 MetGluLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSe 112

337 ATGGAGACGTGAGGAAGGCGAGCTGACCCACAGCCCTCGCGGCGGTAG 288

112 rThrAspProProHisHisAspHisHisAspGluPheCysLeuMetPro 128

287 CACTGACCCGCTCATCATGACCCACCATGATGAGTTTTCCTCATGCC 239

seq\_name: gb\_est2:BF608209

seq\_documentation\_block:

LOCUS BF608209 650 bp mRNA EST 01-APR-2001  
 DEFINITION MVL\_001088 Mouse 9-day fetus cDNA library ICRFP522 Mus musculus  
 cDNA clone ICRFP522C0878 5', mRNA sequence.  
 ACCESSION BF608209  
 VERSION BF608209.1 GI:13504638  
 KEYWORDS EST.

## SOURCE

## ORGANISM

house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 650)

## REFERENCE

## AUTHORS

Yahyawi,M., Hennig,S., Neidhardt,L., Radelof,U., Hermann,B.G.,  
 Lehrach,H. and O'Brien,J.

## TITLE

Detection of a high number of novel genes in a 9-day mouse embryo  
 cDNA library normalised by oligonucleotide fingerprinting

## JOURNAL

## COMMENT

Unpublished (2001)  
 Contact: Hennig S  
 laboraty 123, dept. Lehrach  
 Max-Planck-Institut fuer Molekulare Genetik  
 Ihnestr.63-73, D-14195 Berlin, Germany  
 Tel: +49 30 8413 1612  
 Fax: +49 30 8413 1380  
 Email: hennig@molgen.mpg.de  
 EST's are made from clones being representatives of clone clusters.  
 Clone clusters were calculated from oligonucleotide fingerprints.  
 PCR Primers

FORWARD: 5'-GAGCTATTCCAGAAGTAGTGA-3'

BACKWARD: 5'-TAATAGCACTCACTATAGGG-3'

Seq primer: 5'-ATTAGGTGACACTATAG-3'

High quality sequence stop: 650.

## FEATURES

source

1..650

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="ICRFP522C0878"

/clone\_lib="Mouse 9-day fetus cDNA library ICRFP522"

/tissue\_type="whole embryo"

/dev\_stage="embryonic 9-day"

/lab\_host="E.coli, XL1 blue"

/note="Vector: pSVSport1; Site\_1: NotI; Site\_2: SalI;

Library preparation by oligo\_dT priming of RNA. Clones can  
 be ordered from the Resource Center in Berlin,  
 http://www.rzpd.de."

BASE COUNT 164 a 167 c 180 g 128 t 11 others

ORIGIN

## alignment\_scores:

Quality: 516.00 Length: 128  
 Ratio: 4.526 Gaps: 0  
 Percent Similarity: 89.062 Percent Identity: 75.781

## alignment\_block:

US-09-327-750d-33 x BF608209 ..

Align seg 1/1 to: BF608209 from: 1 to: 650

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHi 17

206 ATGGGACTCCAAAGATCAAGCGGTGAAAAATCTCACATGGACACGACCA 255

17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleLysArgS 34

256 TCACAAAAAGGACGACAGGACAAAGCCACAAAGATACCATCTTAATGG 305

34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50

306 ACCGAGCTGTGGCCCTTACCTCTAGGCTGGCAAAACTGTGCACCTTGA 355

51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyrArgTr 67

356 GNAGGTCCGAGCGGTTCCGGGTTCCGAGCCCATCGCTTACTATANATC 405

67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluAla 84

406 GGACCTGATCCACAGGGTTGGGGAGCCCATGGAGGATCAGAGACGACA 455

84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100

```

456 ACGTACAGAGGCTCTGGGGTGATGTGAGACAGCTCATGAGAAACCTCAGG 505
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProHis 117
|||||
506 GAAAGCAGCTTAGCCACAGCTGCGGGGGTGTAGCACTGACNCGCCTCA 555
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
556 TCATGACTACCATGATTAGTTTGGCTCATGCCC 589
|||||

seq_name: gb_est1:BE291071

seq_documentation_block:
LOCUS BE291071 583 bp mRNA EST 13-JUL-2000
DEFINITION 601086311F1.NC1.CGAP_Mam6 Mus musculus cDNA clone IMAGE:3500522 5',
mRNA sequence.
ACCESSION BE291071
VERSION BE291071.1 GI:9172545
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 583)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
DNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM8559 row: e column: 03
High quality sequence start: 5
High quality sequence stop: 503.
FEATURES
Location/Qualifiers
1..583
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3500522"
/clone_lib="NC1.CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sali;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 165 a 136 c 201 g 81 t
ORIGIN

alignment_scores:
Quality: 505.50 Length: 120
Ratio: 4.554 Gaps: 2
Percent Similarity: 92.500 Percent Identity: 83.333

alignment_block:
US-09-327-750D-33 x BE291071 ..
Align seg 1/1 to: BE291071 from: 1 to: 583
1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHis 17
|||||
208 ATGGAGTCCAAAGATCAAGCGGTGAATAATCTCAACATGGAGGAGCA 257
|||||
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleLysArgG 34
|||||

```

```

258 TCAGAAAAAGGAGGAGGAGAAAGAAAGCCACAGATACCATCAGAAGG 307
|||||
34 luProValValalaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
|||||
308 AGCCAGCTGTGGCCCTGACCTCGAGGCTGGCAAAACCTGTGCACCTAGA 357
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyrArgTr 67
|||||
358 GGAGGTGCGAGCGGTTCCGGGTTCGGACGCCCATCGCTCACTATAGATG 407
|||||
67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGlu 84
|||||
408 GCACCTGATGCAGAGGGTTGGGAGGCCCCAGTCAGAGATGAGAGGACA 457
|||||
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
|||||
458 ACCTACAGAGGTTAGGGGTGATGTGAGACAGCTCATGGAGAAGCTGAGG 507
|||||
101 GluArgGlnLeuSerHis...SerLeuArgAlaVal.SerThrAspProp 116
|||||
508 GAAGGCTGTTGAGCCACAGACATGCGAGCGGTTAGACAATGACCCGA 557
|||||
116 roHisHis 118
|||||
558 CTCATCAT 565

seq_name: gb_gss:AZ936393

seq_documentation_block:
LOCUS AZ936393 637 bp DNA GSS 26-APR-2001
DEFINITION 2M0193L05F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0193L05 F, DNA sequence.
ACCESSION AZ936393
VERSION AZ936393.1 GI:13794974
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 637)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: Q193 row: L column: 05
Seq primer: CGTTGTAAACAGACGGCCACT
Class: plasmid ends
High quality sequence stop: 637.
FEATURES
Location/Qualifiers
1..637
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0193L05"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource

```







**THIS PAGE BLANK (USPTO)**

```
/db_xref="taxon:10090"
/clone="G0106H08"
/clone_lib="NIA Mouse E7.5 Embryonic Portion cDNA Library"
/sex="unknown"
/dev_stage="7.5dpc Embryo"
/lab_host="DH10B"
/Note="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
6 Embryo. The double-stranded cDNA was synthesized by
Gibco's kit with an oligo(dT) primer (NotI primer-adaptor
from GibcoBRL)
[5'-pGACTAGTCTTGTAGATCGAGCGCGCCCTTTTTTTTTTTT-3']
from 0.5ug of mRNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker L1-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Minoru S. H. Ko."
```

BASE COUNT 151 a 151 c 117 g 130 t  
ORIGIN

alignment\_scores:  
Quality: 472.00 Length: 97  
Ratio: 5.187 Gaps: 0  
Percent Similarity: 93.814 Percent Identity: 88.660

alignment\_block:

US-09-327-750D-33 x AW536634/rev ..

Align seg 1/1 to reverse of: AW536634 from: 1 to: 549

```
32 LysArgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAl 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 AAAAGGATCCGATTCTGCCCTGCCCTTCGAAGCTGGAGACTACTACGT 492
48 aProArgGlyGlyArgArgPheArgValArgGlnProIleSerHist 65
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491 GCCTAGAGAGGTTCGAGCGGTTCCGGGTTCCGACGCCCATCGTGCAC 442
65 YrArgTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArg 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 ACAGATGGACCTGATGCATAGGTTGGGAGCCCCCGGAGGATGAGA 392
82 GluGluAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGlu 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||
391 GAGGACAACGTACAGAGTTTGGGGATGATGTCAGACACTCATGAGAA 342
98 sLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 GCTGAGGAAAGGCAGCTGAGCCACACCTGCGGGGTTAGCACCTGACC 292
115 rGProHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 CGCCTCATCATGACCACCATGATGAGTTTTCCTTATGCC 251
```

OM of: US-09-327-750D-33 to: Issued\_Patents\_NA:\* out\_format : pfs

Date: Mar 11, 2002 3:51 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101154\_20340/app\_query.fasta\_1.1472  
-DB=Issued\_Patents\_NA -OPMT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -START=1 -MATRIX=blossum62  
-DELOP=6.000 -DELEXT=7.000 -ALIGN=200 -THR\_SCORE=pct  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=/US09327750.@CGL1\_1\_99 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-33

Query length: 128

Database: Issued\_Patents\_NA:\*

Database sequences: 351203

Database length: 113238999

Search time (sec): 146.090000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-381-881-1 +	93.50	166.92	0.1200	1315	!
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-281-221-1 +	93.50	166.92	0.1200	1315	!
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-07-951-715A-6 +	93.00	155.63	0.5107	3624	!
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-459-448A-6 +	93.00	155.63	0.5107	3624	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-459-595A-6 +	93.00	155.63	0.5107	3624	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-459-504B-6 +	93.00	155.63	0.5107	3624	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-459-44A-6 +	93.00	155.63	0.5107	3624	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-553-549-1 +	93.00	155.63	0.5107	3624	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-053-549-1 +	91.00	142.54	2.74	8854	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-506-553C-6 +	82.50	152.65	0.7483	607	!
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-998-416-994 -	82.00	149.79	1.08	729	!
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-426-599B-3 +	81.00	149.35	1.14	624	!
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-147-777-3 +	80.50	110.90	158.26	24979	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-452-872-3 +	80.50	110.90	158.26	24979	!
/cgn2_6/ptodata/2/ina/PTCTUS.COMB.seq:PCT-US93-03985-3 +	80.50	110.90	158.26	24979	!
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-434-288-10 +	80.00	137.79	5.03	1600	!
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-030-995-1 +	79.00	137.39	5.30	1365	!
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-117-927-4 -	78.50	131.71	10.97	2163	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-931-999-4 +	77.50	118.14	62.53	6755	!
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-125-468-1 -	77.50	103.01	435.42	30001	!
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-474-933-1 -	77.50	103.01	435.42	30001	!
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-998-416-141 -	77.00	140.56	3.53	672	!
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-047-041A-24 -	77.00	133.96	8.23	1288	!
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-285-440-11 +	76.50	130.76	12.40	1599	!
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-630-349-11 +	76.50	130.76	12.40	1599	!
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-285-440-13 +	76.50	127.81	18.09	2137	!
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-630-349-13 +	76.50	127.81	18.09	2137	!
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-306-691B-23 +	76.50	127.06	19.92	2301	!
/cgn2_6/ptodata/2/ina/PTCTUS.COMB.seq:PCT-US93-06251-78 +	76.50	127.06	19.92	2301	!
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-056-200-93 +	76.50	112.62	127.02	9551	!
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-800-644-93 +	76.50	112.62	127.02	9551	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-036-987A-1 -	76.00	90.02	2.3e+03	80161	!
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-370-700-1 -	76.00	90.02	2.3e+03	80161	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-717-294-42 +	75.50	118.36	60.85	4451	!
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-103-840A-2 +	75.50	48.36	3.5e+05	4403765	!
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-381-881-3 +	75.00	137.29	5.37	624	!
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-281-221-3 +	75.00	137.29	5.37	624	!
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-809-740A-1 -	75.00	120.54	46.01	3252	!
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-809-740A-4 -	75.00	120.54	46.01	3252	!
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-822-586-45 +	75.00	109.04	201.00	10095	!

/cgn2\_6/ptodata/2/ina/5B.COMB.seq:US-08-452-075-1 - 74.50 120.90 43.90 2841  
/cgn2\_6/ptodata/2/ina/6B.COMB.seq:US-09-231-061-1 - 74.50 120.90 43.90 2841  
/cgn2\_6/ptodata/2/ina/5B.COMB.seq:US-08-403-852D-1 + 74.50 114.40 101.08 5392  
/cgn2\_6/ptodata/2/ina/6A.COMB.seq:US-08-510-846B-1 + 74.50 114.40 101.08 5392  
/cgn2\_6/ptodata/2/ina/6B.COMB.seq:US-09-231-818-1 + 74.50 114.40 101.08 5392

seq\_name: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:US-08-381-881-1

seq\_documentation\_block:  
; Sequence 1, Application US/08381881  
; Patent No. 5939064  
; GENERAL INFORMATION:  
; APPLICANT: SAVELKOU, PAUL H., M.  
; APPLICANT: GAASTRA, WILLEM  
; TITLE OF INVENTION: BORDETTELLA BRONCHISEPTICA VACCINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STEVENS, DAVIS, MILLER & MOSHER  
; STREET: 515 NORTH WASHINGTON STREET.  
; CITY: ALEXANDRIA  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22314  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/381.881  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: POULOS III, JAMES A.  
; REGISTRATION NUMBER: 31,714  
; REFERENCE/DOCKET NUMBER: TPP 29685  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703 549-7200  
; TELEFAX: 703 528-5313  
; TELEX: 44-0704 "STEVENS"  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1315 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bordetella bronchiseptica  
; STRAIN: 401  
; IMMEDIATE SOURCE:  
; CLONE: E coli PC2495(pIVB3-420)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..539  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 540..1142  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1143..1315  
; US-08-381-881-1

alignment\_scores:  
Quality: 93.50 Length: 96  
Ratio: 1.948 Gaps: 4  
Percent Similarity: 50.000 Percent Identity: 33.333  
alignment\_block:  
US-09-327-750D-33 x US-08-381-881-1

```
Align seg 1/1 to: US-08-381-881-1 from: 1 to: 1315
43 AlaGlyLysAsnCysAlaProArgGlyGlyArgArgPheArgValAr 59
||||:||||| ||| ||||| ||||| |||||
542 GCAAGCGCAAAACGTTCTCTCGCGCGCGCGCTCGCGCGCTCG 591
59 gGlnProIle.....SerHisTyrArgTrrp..... 67
||||| ||||| ||||| |||||
592 CCGCCCATGCCGAGAGCGGCACCATTTGTCATTACCGGCACGATCACCGAC 641
68 ..AspLeuMethHisArgValGlyGluPro.....GlnGlyArgMet 80
||||| ||| ||||| ||||| |||||
642 CAGACCTGCACGATCGAGGACCGCGCGGTTACATCAAGTCTGTGCA 691
81 ArgGluGluAsnValGlnArgPheGlyGlyAspMetArgGlnLeuMetGl 97
||||| ||||| ||||| |||||
692 CCT.....GCCACGATCTCCAGAGCGCGCTGA 720
97 uLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThra 114
||||:||||| ||||| ||||| ||||| |||||
721 AGAAGCGCGCGAGCTGGCGGCGCGACTCGCTTCGATATCAAGCTGAAG 770
114 spProProHisHisAspHisHisAspPheCysLeu 126
||| ||| ||||| |||||
771 GACTGCCCGACACCGCTCAACACTCTCAAGCTGTACTT 808

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-281-221-1

seq_documentation_block:
; Sequence 1, Application US/09281221
; Patent No. 6284256
; GENERAL INFORMATION:
; APPLICANT: GAASTRA, WILLEM
; TITLE OF INVENTION: BORDETELLA BRONCHISEPTICA VACCINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STEVENS, DAVIS, MILLER & MOSHER
; STREET: 515 NORTH WASHINGTON STREET.
; CITY: ALEXANDRIA
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/281,221
; FILING DATE: 30-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/381,881
; ATTORNEY/AGENT INFORMATION:
; NAME: POULOS III, JAMES A.
; REGISTRATION NUMBER: 31,714
; REFERENCE/DOCKET NUMBER: TPP 29685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 549-7200
; TELEFAX: 703 528-5313
; TELEX: 44-0704 "STEVENS"
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```

```
; ORIGINAL SOURCE:
; ORGANISM: Bordetella bronchiseptica
; STRAIN: 401
; IMMEDIATE SOURCE:
; CLONE: E coli PC2495(p1VB3-420)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..539
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 540..1142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1143..1315
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-281-221-1

alignment_scores:
Quality: 93.50 Length: 96
Ratio: 1.948 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 33.333

alignment_block:
US-09-327-750D-33 x US-09-281-221-1 ..
Align seg 1/1 to: US-09-281-221-1 from: 1 to: 1315
43 AlaGlyLysAsnCysAlaProArgGlyGlyArgArgPheArgValAr 59
||||:||||| ||| ||||| ||||| |||||
542 GCAAGCGCAAAACGTTCTCTCGCGCGCGCGCTCGCGCGCTCG 591
59 gGlnProIle.....SerHisTyrArgTrrp..... 67
||||| ||||| ||||| |||||
592 CCGCCCATGCCGAGAGCGGCACCATTTGTCATTACCGGCACGATCACCGAC 641
68 ..AspLeuMethHisArgValGlyGluPro.....GlnGlyArgMet 80
||||| ||| ||||| ||||| |||||
642 CAGACCTGCACGATCGAGGACCGCGCGGTTACATCAAGTCTGTGCA 691
81 ArgGluGluAsnValGlnArgPheGlyGlyAspMetArgGlnLeuMetGl 97
||||| ||||| ||||| |||||
692 CCT.....GCCACGATCTCCAGAGCGCGCTGA 720
97 uLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThra 114
||||:||||| ||||| ||||| ||||| |||||
721 AGAAGCGCGCGAGCTGGCGGCGCGACTCGCTTCGATATCAAGCTGAAG 770
114 spProProHisHisAspHisHisAspPheCysLeu 126
||| ||| ||||| |||||
771 GACTGCCCGACACCGCTCAACACTCTCAAGCTGTACTT 808

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-07-951-715A-6

seq_documentation_block:
; Sequence 6, Application US/07951715A
; Patent No. 5625136
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nallini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
```

```

; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/951,715A
; FILING DATE: 25-SEP-1992
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3621 /product= "Full-length, maize
; OTHER INFORMATION: optimized cryiB"
; OTHER INFORMATION: /note= "Disclosed in Figure 6."
;
; US-07-951-715A-6
;
; alignment_scores:
; Quality: 93.00 Length: 116
; Ratio: 1.550 Gaps: 4
; Percent Similarity: 51.724 Percent Identity: 26.724
;
; alignment_block:
; US-09-327-750D-33 x US-07-951-715A-6 ..
;
; Align seg 1/1 to: US-07-951-715A-6 from: 1 to: 3624
;
; 14 GluAsnAspHisGlnLysLysGluGluLysProGlnAspTh 30
; |||:|||||:|||||:|||||:|||||:|||||:|||||
; 1026 GAACACGAGCCACCGCGGCCACCAACACGATCATCGCCGACCC 1075
;
; 30 rleLysArgGluProValAlaProThrPheGluAlaGlyLysAsnC 47
; |||:|||||:|||||:|||||:|||||:|||||:|||||
; 1076 TGCCTTTCGCGAGCCGCGAGCTGTATCCGACCGAGAGCTACGCCGGCGTG 1125
;
; 47 ysAla.....ProArgGlyGlyArgArg 55
; |||:|||||:|||||:|||||:|||||:|||||:|||||
; 1126 CTGCTGTGGGCATCTACCTGGAGCCCATCATCGGCGTCCCGCGTGGC 1175
;
; 56 PheArgValArgGlnProIleSerHisTyrArgTrpAspLeuMetHisAr 72
; |||:|||||:|||||:|||||:|||||:|||||:|||||
; 1176 CTTCAACTTCCACCAACCCCGGACATCAGCGA.....CCGCGGCACCG 1219

```

```

72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
; |||:|||||:|||||:|||||:|||||:|||||:|||||
1220 CCAACTACAGCCA.....GCCCTACGAGAGCCCGCGCTGCAGCTGAAG 1263
;
89 lyGluAspMetArgGlnLeuMetGluLysLeuArgGlnLeuSer 105
; |||:|||||:|||||:|||||:|||||:|||||:|||||
1264 GACAGCGAGACCGCTGCCCGCCGAGACACCGAGCGGCCCAA..... 1307
;
106 HisSerLeuArgAlaValSerThrAspProProHisHisAspHis 121
; |||:|||||:|||||:|||||:|||||:|||||:|||||
1308 .....CTACGAGAGCTACGACCGCCCTGTAGCCACATCGGCATCAT 1349
;
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-459-448A-6
;
seq_documentation_block:
; Sequence 6, Application US/08459448A
; Patent No. 5859336
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5859336artis Corporation
; STREET: Patent & Trademark Dept., 520 White Plains
; STREET: Rd., POB 2005
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-9005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,448A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/951,715
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40403
; REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8582
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: 1..3621
; OTHER INFORMATION: /product= "Full-length, maize
; OTHER INFORMATION: optimized cryIB"
; OTHER INFORMATION: /note= "Disclosed in Figure 6."
;
US-08-459-448A-6

alignment_scores:
  Quality: 93.00 Length: 116
  Ratio: 1.550 Gaps: 4
  Percent Similarity: 51.724 Percent Identity: 26.724

alignment_block:
US-09-327-750D-33 x US-08-459-448A-6 ..
  Align seg 1/1 to: US-08-459-448A-6 from: 1 to: 3624

14 GluAsnAspHisGlnLysLysGluGluLysGluLysProGlnAspTh 30
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1026 GAACACGACGACCCACCGCGGCCCAACACGACGATCAACCCCGTGACCC 1075

30 rIleLysArgGluProValValAlaProThrPheGluAlaGlyLysAsnC 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||
1076 TGCCTTCGCGACGCGCGGCGGTGTACCGACCGGAGAGCTACCGCGCGTG 1125

47 ysAla.....ProArgGlyGlyArgArg 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||
1126 CTGCTGTGGGCGATCTACCTGGAGCCCATCCAGCGCTGCCACCGTGCG 1175

56 PheArgValArgGlnProIleSerHisTyrArgTrpAspLeuMethHisAr 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||
1176 CTTCAACTTCACCAACCCCGACGACATCAGCGA.....CCGCGGCACCG 1219

72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||
1220 CCAACTACAGCA.....GCCCTACGAGAGCGCGCCCTGCAGCTGAAG 1263

89 lyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSer 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||
1264 GACAGCGAGACGAGCTGCCCGCCGAGACCAACGAGCGGCCCAAA..... 1307

106 HisSerLeuArgAlaValSerThrAspProProHisHisAspHisHis 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||
1308 .....CTACGAGAGCTACAGCCACCGCGCTGAGCCACATCGGCATCAT 1349

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-595A-6

seq_documentation_block:
; Sequence 6, Application US/08459595A
; Patent No. 6018104
;
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Meriin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
```



```
56 PheArgValArgGlnProHisSerHisTyrArgTrpAspLeuMethHisAr 72
1176 CTTCACTTCCACCAACCCCGAGAACATCAGCGA.....CCGCGGCACCG 1219
72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
1220 CCAACTACAGCCA.....GCCCTACGAGAGAGCCCGCGCTGCAGCTGAAG 1263
89 lyGluAspMetArgGlnLeuMetGlnLysLeuArgGluArgGlnLeuSer 105
1264 GACAGCGAGACCGAGCTGCCCGGAGACACCGAGCGGCCCCAA..... 1307
106 HisSerLeuArgAlaValSerThrAspProHisHisAspHisHis 121
1308 .....CTACGAGAGCTACAGCAGCGCTGAGCCACATCGGCATCAT 1349
```

seq\_name: /cqn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-459-504B-6

```
seq_documentation_block:
; Sequence 6, Application US/08459504B
; Patent No. 6075185
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6075185artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,504B
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/459,595
; FILING DATE: 02-JUN-1995
; APPLICATION NUMBER: US 07/951,715
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1577/CIP/DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8587
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 6:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3621
; OTHER INFORMATION: /product= "Full-length, maize
; OTHER INFORMATION: optimized cryIB"
; OTHER INFORMATION: /note= "Disclosed in Figure 6."
US-08-459-504B-6

alignment_scores:
Quality: 93.00 Length: 116
Ratio: 1.550 Gaps: 4
Percent Similarity: 51.724 Percent Identity: 26.724

alignment_block:
US-09-327-750D-33 x US-08-459-504B-6 ..
Align seg 1/1 to: US-08-459-504B-6 from: 1 to: 3624

14 GluAsnAspHisGlnLysLysGluGluLysGluGluLysProGlnAspTh 30
||||:||||| :||||:||||| :||||:||||| :||||:|||||
1026 GAACACGACGACCGCGCGCCACCAACACGACGATCAACCGCGTGACCC 1075
30 rleLysArgGluProValValAlaProThrPheGluAlaGlyLysAsnC 47
||||:||||| :||||:||||| :||||:||||| :||||:|||||
1076 TGGCGCTTGGCGGCGGAGCTGTACCGACGAGAGCTAGCGCGCGGTG 1125
47 ysAla.....ProArgGlyGlyArgArg 55
||||:||||| :||||:||||| :||||:||||| :||||:|||||
1126 CTGCTGTGGGCGCATCTACCTGGAGCCCATCCACGCGTGCCACCGTGC 1175
56 PheArgValArgGlnProHisSerHisTyrArgTrpAspLeuMethHisAr 72
||||:||||| :||||:||||| :||||:||||| :||||:|||||
1176 CTTCACTTCCACCAACCCCGAGAACATCAGCGA.....CCGCGGCACCG 1219
72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
||||:||||| :||||:||||| :||||:||||| :||||:|||||
1220 CCAACTACAGCCA.....GCCCTACGAGAGAGCCCGCGCTGCAGCTGAAG 1263
89 lyGluAspMetArgGlnLeuMetGlnLysLeuArgGluArgGlnLeuSer 105
||||:||||| :||||:||||| :||||:||||| :||||:|||||
1264 GACAGCGAGACCGAGCTGCCCGGAGACACCGAGCGGCCCCAA..... 1307
106 HisSerLeuArgAlaValSerThrAspProHisHisAspHisHis 121
||||:||||| :||||:||||| :||||:||||| :||||:|||||
1308 .....CTACGAGAGCTACAGCAGCGCTGAGCCACATCGGCATCAT 1349

seq_name: /cqn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-444-6

seq_documentation_block:
; Sequence 6, Application US/08459444A
; Patent No. 6121014
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED
; NUCLEIC ACID CODING SEQUENCE
```

```

;
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6121014artis Agribusiness Biotechnology Research, Inc.
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,444A
; FILING DATE: 02-Jun-1995
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/951,715
; FILING DATE: 25-SEP-1992
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: S-18805/PI/CGC1577/CIP/DIV6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8587
; TELEFAX: (919)541-8689
;
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: 1..3621
; OTHER INFORMATION: /product= "Full-length, maize
; optimized cryiB"
; /note= "Disclosed in Figure 6."
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
;
; US-08-459-444-6
;
;
; alignment_scores:
; Quality: 93.00 Length: 116
; Ratio: 1.550 Gaps: 4
; Percent Similarity: 51.724 Percent Identity: 26.724
;
; alignment_block:
; US-09-327-750D-33 x US-08-459-444-6 ..
;
; Align seg 1/1 to: US-08-459-444-6 from: 1 to: 3624
;
; 14 GluAsnAspHisGlnLysLysGluGluLysProGlnAspHs 30
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 1026 GAACACGAGCAGCCAGCGCGCCACCAACAGCAGCATCAACCGGTGACCC 1075
;
; 30 rLeLysArgGluProValAlaProThrPheGluAlaGlyLysAsnC 47
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 1076 TGCCTTCGCGAGCGCGGACGTGTACCGCACCGAGAGCTACGCGCGGTG 1125
;
; 47 ysAla.....ProArgGlyGlyArgArgArg 55
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 1126 CTGCTGTGGGCGATCTACCTGGAGCCCATCCACGCGGTGCCACCGTGGC 1175
;
; 56 PheArgValArgGlnProIleSerHisTyrArgTrpAspLeuMethIsAr 72
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 1176 CTTCAACTTCACCAACCCCAAGAACATCAGCGA.....CCGCGGACCG 1219
;

```

```

72 gValGlyGluProGlnArgMetArgGluGluAsnValGlnArgPheG 89
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 1220 CCAACTACACCA.....GCCCTACGAGAGAGCCCGCGCTGCAGCTGAAG 1263
;
; 89 lyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSer 105
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 1264 GACAGCGAGACCGAGCTGCCCGCCGAGACCGAGCGGCCCA..... 1307
;
; 106 HisSerLeuArgAlaValSerThrAspProPheHisHisAspHis 121
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 1308 .....CTAGGAGAGCTACAGCCACCGCTGAGCCACATCGGCATCAT 1349
;
; seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-053-549-7
;
; seq_documentation_block:
; Sequence 7, Application US/09053549
; Patent No. 6121521
; GENERAL INFORMATION:
; APPLICANT: Desai, Nallini
; TITLE OF INVENTION: No. 6121521el Insecticidal Protein and Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6121521artis Corporation
; STREET: 3054 Cornwallis Rd.
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053,549
; FILING DATE: 01-APR-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1995
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
;
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: 1..3621
; OTHER INFORMATION: /product= "Full-length, maize
; optimized cryiB"
; OTHER INFORMATION: /note= "Disclosed in Figure 6."
;
; US-09-053-549-7
;

```

```

; alignment_scores:
; Quality: 93.00 Length: 116
; Ratio: 1.550 Gaps: 4
; Percent Similarity: 51.724 Percent Identity: 26.724
;
; alignment_block:
; US-09-327-750D-33 x US-09-053-549-7 ..
;
; Align seg 1/1 to: US-09-053-549-7 from: 1 to: 3624
;

```

```
14 GluAsnAspHisGlnLysLysGluGluLysGluLysProGlnAspTh 30
||||:||||| :||||:||||| :||||:|||||
1026 GAACACGAGCAGCCAGCGGCCACCAACAGCATCAACCCCGTGACCC 1075
30 rleLysArgGluProValValAlaProThrPheGluAlaGlyLysAsnC 47
||||:||||| :||||| :||| :||| :|||
1076 TGGCGCTTGGCGAGCGCGAGCGTGTACCGACCGAGAGCTACGCGCGGTG 1125
47 ysAla.....ProArgGlyGlyArgArgArg 55
:||||| :||||| :||||| :|||||
1126 CTGCTGTGGGGCATCTACCTGGAGCCCATCCAGCGGTGCCACCGTGGC 1175
56 PheArgValArgGlnProIleSerHisTyrArgTirpAspLeuMethHisAr 72
:||||:||||| :||||| :||| :||| :|||
1176 CTTCAACTTCACCAACCCCGCAGAACATCAGCGA.....CCGCGGCACCG 1219
72 gvalGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
| :||||| :||||| :||| :||| :|||
1220 CCAACTACAGCCA.....GCCCTACGAGAGAGCCCGCGCTGCAGCTGAAG 1263
89 lyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSer 105
||||| :||| :||| :||| :|||
1264 GACAGCAGCAGCGAGCTGCCCGCGAGACACCGAGCGGCCCAA..... 1307
106 HisSerLeuArgAlaValSerThrAspProHisHisAspHisHis 121
||||| :||||| :||| :||| :|||
1308 .....CTACGAGAGCTACAGCCAGCCGCTGAGCCACATCGGCATCAT 1349
seq_name: /cgn2.6/ptodata/2/ina/6A_COMB.seq:US-09-053-549-1
```

```
seq_documentation_block:
; Sequence 1, Application US/09053549
; Patent No. 6121521
; GENERAL INFORMATION:
; APPLICANT: Desai, Nalini
; TITLE OF INVENTION: No. 6121521el Insecticidal Protein and Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6121521artis Corporation
; STREET: 3054 Cornwallis Rd.
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053,549
; FILING DATE: 01-APR-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1995
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 11..3694
; OTHER INFORMATION: /product= "hyFLIB protein"
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: 8854
; OTHER INFORMATION: /note= "Sequence of pCIB5520
; Patent No. 6121521
; OTHER INFORMATION: containing coding sequence for hyFLIB protein"
US-09-053-549-1

alignment_scores:
Quality: 91.00 Length: 116
Ratio: 1.542 Gaps: 4
Percent Similarity: 50.862 Percent Identity: 27.586

alignment_block:
US-09-327-750D-33 x US-09-053-549-1 ..

Align seg 1/1 to: US-09-053-549-1 from: 1 to: 8854

14 GluAsnAspHisGlnLysLysGluGluLysGluLysProGlnAspTh 30
||||:||||| :||||:||||| :||||:|||||
1099 GAACACGAGCAGCCAGCGGCCACCAACAGCATCAACCCCGTGACCC 1148
30 rleLysArgGluProValValAlaProThrPheGluAlaGlyLysAsnC 47
||||:||||| :||||| :||| :||| :|||
1149 TGGCGCTTGGCGTCCCGAGAGCTGTACCGCAGCAGAGCTACGCGCGGTG 1198
47 ysAla.....ProArgGlyGlyArgArgArg 55
:||||| :||||| :||||| :|||||
1199 CTGCTGTGGGGCATCTACCTGGAGCCCATCCAGCGGTGCCACCGTGGC 1248
56 PheArgValArgGlnProIleSerHisTyrArgTirpAspLeuMethHisAr 72
:||||:||||| :||||| :||| :||| :|||
1249 CTTCAACTTCACCAACCCCGCAGAACATCAGCGA.....CCGCGGCACCG 1292
72 gvalGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
| :||||| :||||| :||| :||| :|||
1293 CCAACTACAGCCA.....GCCCTACGAGAGCGCGGTTCAGCTGAAG 1336
89 lyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSer 105
||||| :||| :||| :||| :|||
1337 GACAGCAGCAGCAGCTGCCCGCGAGACACCGAGCGGCCCAA..... 1380
106 HisSerLeuArgAlaValSerThrAspProHisHisAspHisHis 121
||||| :||||| :||| :||| :|||
1381 .....CTACGAGAGCTACAGCCAGCCGCTGAGCCACATCGGCATCAT 1422
seq_name: /cgn2.6/ptodata/2/ina/6A_COMB.seq:US-08-506-553C-6

seq_documentation_block:
; Sequence 6, Application US/08506553C
; Patent No. 6120989
; GENERAL INFORMATION:
; APPLICANT: Vornhagen, Rolf; Hinderer, Walter; Sonnenborn, Han-H.;
; APPLICANT: Plachter, Bodo; and Jahn, Gerhard
; TITLE OF INVENTION: ISOLATED HUMAN CYTOMEGALOVIRUS
; TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,553C
; FILING DATE: 07/25/95
; CLASSIFICATION: 435
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LEADER 202-PFF/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-506-553C-6

alignment_scores:
  Quality: 82.50      Length: 117
  Ratio: 1.650       Gaps: 4
  Percent Similarity: 42.735      Percent Identity: 25.641

alignment_block:
US-09-327-750D-33 x US-08-506-553C-6 ..
Align seg 1/1 to: US-08-506-553C-6 from: 1 to: 607

50 ArgGlyGlyArgArgArgPheArgValArgGlnProIleSerHisTyrAr 66
   |||:|||||
107 CGGCTGGGAGAGGTGGCGCGGTGGCGGCAAGAA...GCACGACCGG 153
   |||:|||||

66 gTrpAspLeuMethHisArgValGlyGluProGln..... 77
   |||:|||||
154 GTGGCGGCGGTGGTTCGGTACGGGAAATGAGTAGCGGTGGCGGCGC 203
   |||:|||||
77 ..... 77

204 GGTGATCATGACCGGTCTTCTCCAGGAAAAATACGAGCAGCACAA 253
78 .....GlyArgMetArgGlu1 83
   |||:|||||
254 GATCACCAGCTACCTGAGTCCAAAGGTGGATCGGCGCGCGGAGGAG 303
   |||:|||||
83 uAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuA 100
   |||:|||||
304 GAGGAGCGCGGTTGGATCGCAACTCCGCAATTACTTCAACGACGCG 353
   |||:|||||
100 rg.....GluArgGlnLeuSer..... 105
   |||:|||||
354 AAAGAGGAGCGAGCAGCGAGGATTCTGTACGTTCCGTTCCCTAA 403
   |||:|||||
106 HisSerLeuArgAlaValSerThrAspProHisHisAspHisAs 122
   |||:|||||
404 CACCAAGAGCAAAAGTGGCGCAAGATCTTCAAGAGCAGCAGCGCATGA 453
   |||:|||||
122 p 122
454 A 454

seq_name: /cgn2_6/ptodata/2/ina/68_COMB.seq:US-08-998-416-994
seq_documentation_block:
; Sequence 994, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen

; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 994:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1610UP
; US-08-998-416-994

alignment_scores:
  Quality: 82.00      Length: 112
  Ratio: 1.302       Gaps: 8
  Percent Similarity: 56.250      Percent Identity: 31.250

alignment_block:
US-09-327-750D-33 x US-08-998-416-994/rev ..
Align seg 1/1 to reverse of: US-08-998-416-994 from: 1 to: 729

16 AspHisGlnLysLysGluGluLysGluLysProGlnAspThrIlely 32
   |||:|||||
670 AACCCAGAAAGCTCCACAGCAACGTCGTCGCCAGTT..... 632
   |||:|||||
32 sArgGluProValAlaProThrPheGluAlaGlyLysAsnCysAlap 49
   |||:|||||
631 ....GAACCAAGTACATATACCGCTTTGCCAC.....GCCA 601
   |||:|||||
49 roArg GlyGlyArgArgPhe.ArgValArgGlnProIleSerHisT 65
   |||:|||||
600 AACGAGGCGGACAGAAACCGCTTTGAAATACCGTCACCGAGAGTCAC. 552
   |||:|||||
65 yrArgTrpAspLeuMethHisArgValGlyGluProGlnGlyArgMetArg 81
   |||:|||||
551 ..AGGAATGAACGTCGCCATCTC.....CCCCAG...CGACTGCGT 516
   |||:|||||
82 GluGluAsnValGlnArgPheGlyGluAspMetArgGlnLeu..... 95
   |||:|||||
515 GATCGAGTACACCGGTCACCGATGAAGTGTACGCCAACAGCAGCCCA 466
   |||:|||||
```

```
96 .MetGluLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValis 112
   :::::||||| :::::||||| :::::||||| :::::|||||
465 CTGGCCCGCTGTCGCGGGGATGCAATGACATGACAGCGCTGTG 416
   :::::||||| :::::||||| :::::||||| :::::|||||
112 erThrAspProHisHisAspHisAsp 122
   :::::||||| :::::||||| :::::||||| :::::|||||
415 CCTATGCTCCGCCGCGCATCATGAGAT 384
```

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-426-599B-3

```
seq_documentation_block:
; Sequence 3, Application US/08426599B
; Patent No. 5932438
; GENERAL INFORMATION:
; APPLICANT: Uriach-Marsal, Juan
; APPLICANT: Rubio-Susan, Victor
; APPLICANT: Patino-Martin, Cristina
; APPLICANT: Kalo-Koenova, Eliza I.
; APPLICANT: del Moral-Juarez, Catalina
; APPLICANT: Paus-Santassana, Ignacio
; APPLICANT: del Rio-Pericacho, Jose-Luis
; APPLICANT: Blade-Pique, Joan
; TITLE OF INVENTION: Preparation of Thaumatin Sweeteners
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701 East
; CITY: Tower
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426.599B
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Stephen A.
; REGISTRATION NUMBER: 38,609
; REFERENCE/DOCKET NUMBER: 1604-123A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-624-1589
; TELEFAX: 202-783-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Optimized cDNA"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..621
US-08-426-599B-3
```

alignment\_scores:  
Quality: 81.00 Length: 77  
Ratio: 1.976 Gaps: 4  
Percent Similarity: 53.247 Percent Identity: 36.364

alignment\_block:  
US-09-327-750D-33 x US-08-426-599B-3 ..

Align seg 1/1 to: US-08-426-599B-3 from: 1 to: 624

```
52 GJYArqArgPheArgValArgGlnProIleSerHisTyrArgTrpAs 68
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 GGGCGAGCGCGCCCTCGAGCGCGCGCGAGCTCAACTCCGG..... 101
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 pLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGluAsnV 85
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
102 .....CGAGTCTGTGACCATCACTCGAGCGCGCGACCAAGGCG 141
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 aIcInArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArgGlu 101
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
142 GGCAAGATCTGGCGCGCAC...CGACTGCTACTCGAGGACTCCGCGCG 188
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 ArgGlnLeuSerHis.....SerLe 108
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
189 CGCATCTCGCGACCGGCGGACTGCGCGCGCTCTCCAGTCCAACGCT 238
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
108 uArgAlaValSerThrAspProHisHis 118
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
239 TCGG.....CGCGCGCGCGCAC 257
   ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-147-777-3

seq_documentation_block:
; Sequence 3, Application US/08147777
; Patent No. 5914265
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147.777
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA: Including application two
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/876,289
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: Unassigned (204/144)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Watburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 204/153
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24979 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-147-777-3
```

```
alignment_scores:
  Quality: 80.50      Length: 72
  Ratio: 1.872       Gaps: 1
  Percent Similarity: 59.722  Percent Identity: 31.944

alignment_block:
US-09-327-750D-33 x US-08-147-777-3 ..
Align seg 1/1 to: US-08-147-777-3 from: 1 to: 24979

50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
||||| |||||:||||:||||:||||: |||
5088 CGCGGGAGCGCGCGCAGCTACGGCTCAGGAGCTACGGCGCGCGCAGCG 5137

66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
||||| ||| :||| ||||| ||| :|||
5138 GTGGGGGCTATGGCGCGGGAAGAAGCTACCGCGGAGCGGCGCAGCAGGC 5187

83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
:||||:||||:||||: |||:||||: |||:||||: |||:||||: |||
5188 GGGAGTGGAGCGGGTTATGCGAGCGGCTGCGCGCGGCGGTGGCGGAGCTA 5237

100 ...ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5238 CGGAGGCGCGCAGAGAGCGCGCGCGGATCCTCGCGCGTGCAGATCATCC 5287

115 oProHisHisAspHis 120
| ||||| |||||
5288 AGACCTCCACCAACAC 5303

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-452-872-3

seq_documentation_block:
; Sequence 3, Application US/08452872
; Patent No. 6057298
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452.872
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,777
; FILING DATE:
; APPLICATION NUMBER: 07/876,289
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: Unassigned (204/144)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 204/153
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24979 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-452-872-3

alignment_scores:
  Quality: 80.50      Length: 72
  Ratio: 1.872       Gaps: 1
  Percent Similarity: 59.722  Percent Identity: 31.944

alignment_block:
US-09-327-750D-33 x US-08-452-872-3 ..
Align seg 1/1 to: US-08-452-872-3 from: 1 to: 24979

50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
||||| |||||:||||:||||:||||: |||
5088 CGCGGGAGCGCGCGCAGCTACGGCTCAGGAGCTACGGCGCGCGCAGCG 5137

66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
||||| ||| :||| ||||| ||| :|||
5138 GTGGGGGCTATGGCGCGGGAAGAAGCTACCGCGGAGCGGCGCAGCAGGC 5187

83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
:||||:||||:||||: |||:||||: |||:||||: |||:||||: |||
5188 GGGAGTGGAGCGGGTTATGCGAGCGGCTGCGCGCGGCGGTGGCGGAGCTA 5237

100 ...ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5238 CGGAGGCGCGCAGAGAGCGCGCGGATCCTCGCGCGTGCAGATCATCC 5287

115 oProHisHisAspHis 120
| ||||| |||||
5288 AGACCTCCACCAACAC 5303

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-03985-3

seq_documentation_block:
; Sequence 3, Application PCT/US9303985
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: DEVELOPMENT OF A VECTOR TO TARGET GENE
; TITLE OF INVENTION: EXPRESSION TO THE EPIDERMIS OF TRANSGENIC ANIMALS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03985
; FILING DATE: 19930428
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
```

**THIS PAGE BLANK (USPTO)**

```
/
/ REGISTRATION NUMBER: 32,714
/ REFERENCE/DOCKET NUMBER: D-5478
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 713/651-5325
/ TELEFAX: 713/651-5246
/ TELEX: 762829
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 24979 base pairs
/ TYPE: NUCLEIC ACID
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ PCT-US93-03985-3

alignment_scores:
  Quality: 80.50      Length: 72
  Ratio: 1.872      Gaps: 1
  Percent Similarity: 59.722      Percent Identity: 31.944

alignment_block:
US-09-327-750D-33 x PCT-US93-03985-3  ..
Align seg 1/1 to: PCT-US93-03985-3 from: 1 to: 24979
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
||||| |||||:||||:||||: ||
5088 CGCGGGAGCGCGCGGCGAGCTACGGCTCAGGAGCTACGGCGCGGCGGCG 5137
66 qTrpAspLeuMethHisArgValGlyGluProGlnGlyArgMetArgGluG 83
|||| ||| :||| ||||| ||| :|||
5138 GTGGGGCTATGGCGCGCGGAGAGAGCTACCGCGGCGCGGCGGCGGCGG 5187
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
:|||| :|||: ||| :|||: |||
5188 GGGAGTGGAGCGGTTATGGCAGCGGCTGCGCGCGCGTGGCGGAGCTA 5237
100 ...ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
||||| |||||
5238 CGGAGGAGCGGCGGAGAGCGCGCGGATCTCGCGGTGCAGATCATCC 5287
115 oProHisHisAspHis 120
| |||:||||
5288 AGACCTCCACCAACAC 5303
```



OM of: US-09-327-750D-33 to: N\_Geneseq\_1101.\* out\_format : pfs  
 Date: Mar 11, 2002 3:42 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

# Command line parameters:

-MODEL=framet\_p2n.model -DEV=xlh  
 -O=/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101154\_20362/app\_query.fasta\_1.1472  
 -DB=N\_Geneseq\_1101 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=20000000000  
 -USR=US09327750.@CGL1\_1330 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

# Search information block:

Query: US-09-327-750D-33  
 Query length: 128  
 Database: N\_Geneseq\_1101.\*  
 Database sequences: 930621  
 Database length: 428662619  
 Search time (sec): 355.560000

# score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	468.00	903.97	767	
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	468.00	902.84	862	
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	468.00	902.45	898	
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	466.50	900.72	792	
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	455.00	879.92	662	
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	455.00	879.92	662	
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	455.00	877.41	858	
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	444.50	858.92	692	
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	187.50	352.90	1.4e-11	891
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	173.00	326.82	4.0e-10	700
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	170.00	318.90	1.1e-09	865
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	170.00	315.51	1.7e-09	1229
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	170.00	314.51	1.9e-09	1364
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	158.50	295.80	2.1e-08	917
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	135.50	255.99	3.5e-06	532
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	108.50	203.98	0.0028	485
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	106.50	201.42	0.0038	421
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	106.50	200.63	0.0042	457
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	100.50	197.49	0.0063	187
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	96.00	180.18	0.0584	451
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	93.50	169.24	0.2377	844
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	88.50	166.04	0.3584	426
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	86.00	146.63	4.32	1917
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	86.00	145.40	5.06	2178
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	85.00	144.15	5.94	2024
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	85.00	137.85	13.31	3886
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	85.00	136.20	16.45	4612
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	85.00	136.14	16.59	4642
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	85.00	135.99	16.91	4716
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	84.00	153.43	1.91	631
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	83.50	128.60	43.64	7481
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	83.00	140.05	10.04	2061
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	82.50	150.87	2.51	607
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	82.00	143.74	6.26	1148
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	82.00	132.48	26.52	3688
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	81.00	160.43	0.7355	166
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	81.00	149.76	2.89	502
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	81.00	147.66	3.79	624
/SID22/gcgdata/geneseq/NA2001.DAT:AAH75810	+	81.00	145.40	5.06	789

/SID22/gcgdata/geneseq/NA1995.DAT:AAQ86694 - 81.00 139.01 11.48  
 /SID22/gcgdata/geneseq/NA1995.DAT:AAQ86693 - 81.00 131.00 32.05  
 /SID22/gcgdata/geneseq/NA1992.DAT:AAQ28987 + 80.50 134.81 19.68  
 /SID22/gcgdata/geneseq/NA2000.DAT:AAZ43769 + 80.50 134.51 20.45  
 /SID22/gcgdata/geneseq/NA1999.DAT:AAH75810 + 80.50 111.09 412.18

seq\_name: /SID22/gcgdata/geneseq/NA2001.DAT:AAH75810

seq\_documentation\_block:

ID: AAH75810 standard; CDNA: 767 BP.

XX AAH75810;

DT 17-OCT-2001 (first entry)

DE Human X chromosome linked gene expression protein 14 coding sequence.

KW Human; X chromosome linked gene expression protein 14; cancer;

KW HIV infection; cytostatic; anti-HIV; chromosome X; ss.

OS Homo sapiens.

PN CN1296969-A.

XX 30-MAY-2001.

XX 23-NOV-1999; 99CN-0124078.

PR 23-NOV-1999; 99CN-0124078.

XX (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.

XX Mao Y, Xie v;

DR WPT; 2001-483897/53.

DR P-PSDB; AAG66407.

PT Polypeptide-human X chromosome linked gene expression protein 14 and polynucleotide for coding said polypeptide -

PS Claim 6; Page 24 (disclosure); 31pp; Chinese.

CC The present sequence is the coding sequence for human X chromosome linked gene expression protein 14. The protein and coding sequence are useful for treating diseases e.g. cancer and HIV infection.

XX Sequence 767 BP; 224 A; 143 C; 215 G; 185 T; 0 other;

# alignment\_scores:

Quality: 468.00 Length: 129  
 Ratio: 4.000 Gaps: 2  
 Percent Similarity: 90.698 Percent Identity: 68.217

# alignment\_block:

US-09-327-750D-33 x AAH75810

Align seg 1/1 to: AAH75810 from: 1 to: 767

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16  
 |||||  
 101 ATGGAGTCCAAAGAGACGCGTTAAACATCTCATCGTGGAATGCT 150  
 16 phisGlnLysLysGluGluLysGluLysProGlnAspThrIleLysA 33  
 |||||  
 151 CAACAGGAGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 200  
 33 rglGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49  
 |||||  
 201 GGAGGCCCTTGGCCCTTACCT...TTGAATGTTAGTAATCTGTGTGCT 247  
 50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyAr 66  
 |||||

248 AGAGAAACCGTAGCGCGTTCCGCGTTAGGAGCCCATCTCTGAGTATAG 297  
66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
298 ATGGGACATATGTCATAGGCTTGAGAGCCACAGGCAAGCATGAGAGAG 347  
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
348 AGAATATGGAAGGATTGGGAGAGGTGAGACAGCTGATGGAAAGCTG 397  
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
398 AGGAAAAGCAGTTGACTCATAGTCTCGCGGCAGTCAGCACTGATCCCC 447  
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
448 TCACCATGACCATCAGCATGAGTTTGCCTTATGCCCC 484

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI58581

seq\_documentation\_block:  
ID AAI58581 standard; cDNA; 862 BP.

XX AC AAI58581;  
XX DT 22-OCT-2001 (first entry)  
XX DE Human polynucleotide SEQ ID NO 784.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;  
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX KW leukaemia; ss.

XX OS Homo sapiens.  
XX PN WO200153312-A1.  
XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.  
XX PR 21-JAN-2000; 2000US-0488725.  
XX PR 25-APR-2000; 2000US-0552317.  
XX PR 09-JUL-2000; 2000US-0598042.  
XX PR 19-JUL-2000; 2000US-0620312.  
XX PR 03-AUG-2000; 2000US-0653450.  
XX PR 14-SEP-2000; 2000US-0662191.  
XX PR 19-OCT-2000; 2000US-0693036.  
XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;  
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI; 2001-442253/47.  
XX DR P-PSDB; AAM39425.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders  
XX PT such as central nervous system injuries -

XX PS Claim 1; SEQ ID NO 784; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide  
XX CC of the invention may be used to treat diseases of the peripheral nervous  
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification.

SQ Sequence 862 BP; 235 A; 183 C; 238 G; 206 T; 0 other;

alignment\_scores:  
Quality: 468.00 Length: 129  
Ratio: 4.000 Gaps: 2  
Percent Similarity: 90.698 Percent Identity: 68.217

alignment\_block:  
US-09-327-750D-33 x AAI58581 ..

Align seg 1/1 to: AAI58581 from: 1 to: 862

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
226 ATGGAGTCCAAAGAGGAGGAGCGGTTAAACAATCTCATCTGCGAAATGT 275  
16 pHisGlnLysLysGluGluLysGlnGluLysProGlnAspThrIleLysA 33  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
276 CAACCAAGGAAATGATGAAAAAGATGAAAAGGAGCAAGTTGCTAATAAAG 325  
33 rGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
326 GGGAGCCCTTGGCCCTTACCT...TTGAATGTTAGTGAATACTGTGTGCT 372  
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyraR 66  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
373 AGAGGAAACCGTAGCGGTTCCGCGTTAGGAGCCCATCTCTGAGTATAG 422  
66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
423 ATGGGACATAATGCTAGGCTTGGAGAGCCACAGGCAAGCATGAGAGAG 472  
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
473 AGAATATGGAAGGATTGGGAGAGGAGTGAGAGCTGATGGAAAGCTG 522  
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
523 AGGAAAAGCAGTTGAGTCTAGTTTGGCGGCGAGTCAGCACTGATCCCC 572  
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
573 TCACCATGACCATCAGCATGAGTTTGCCTTATGCCCC 609

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF59611

seq\_documentation\_block:

ID AAF59611 standard; cDNA; 898 BP.

XX AC AAF59611;

XX DT 24-APR-2001 (first entry)

XX DE Human cell cycle and proliferation protein CCYPR-22 cDNA, SEQ ID NO:76.  
XX KW Cell cycle and proliferation protein; CCYPR; human; agonist;  
XX KW antagonist; gene therapy; detection; gene therapy;  
XX KW transgenic animal disease model; immune disorder;  
XX KW developmental disorder; cell signalling disorder;  
XX KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
XX KW arteriosclerosis; asthma; allergy; diabetes mellitus;

```

menstrual cycle disorder; bacterial infection; ss.
XX
XX Homo sapiens.
WO200107471-A2.
XX
XX 01-FEB-2001.
PD
XX
XX 21-JUL-2000; 2000WO-US19948.
PF
XX
XX 21-JUL-1999; 99US-0145075.
PR
XX 08-SEP-1999; 99US-0153129.
PR
XX 10-NOV-1999; 99US-0164647.
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
PI Azinzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
PI
XX WPI; 2001-112727/12.
DR
XX P-PSDB; AAB60474.
XX
XX Human cell cycle and proliferation proteins and polynucleotides are
PT used to treat, diagnose and prevent immune, developmental and cell
PT signaling disorders and cell proliferative disorders including cancer -
XX Claim 5; Page 181-182; 205pp; English.
XX
XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC associated with decreased expression of functional CCYPR, while CCYPR
CC antagonists are used to treat diseases or conditions associated with
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCYPR, and in drug screening methods to
CC identify compounds that modulate the activity of CCYPR. CCYPR
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCYPR for the
CC treatment or prevention of a disorder associated with CCYPR.
CC Diseases which can be diagnosed, treated and prevented using CCYPR
CC proteins, nucleic acids, agonists or antagonists include immune,
CC developmental and cell signalling disorders, and cell proliferative
CC disorders including cancer. Specific examples of these disorders
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections
CC caused by bacteria.
XX
SQ Sequence 898 BP; 250 A; 186 G; 251 G; 211 T; 0 other;

alignment_scores:
Quality: 468.00 Length: 129
Ratio: 4.000 Gaps: 2
Percent Similarity: 90.698 Percent Identity: 68.217

alignment_block:
US-09-327-750D-33 x AAF59611 ..
Align seg 1/1 to: AAF59611 from: 1 to: 898

1 MetGluSerTysAspGlnGlyala...LysAsnLeuAsnMetGluAAsAs 16
||||| ||||| : : : : : ||| : : : : : |||
254 ATGAGTGTCCTCAAGAGACGCGTAAACAATCTCATCTGGAAAATGT 303

16 pHtGlnLysLysGluGluScluLysProGlnAspThrIleLysA 33
||||| : : : : : ||||| : : : : : ||| : : : : : |||
304 CAACAGGAANAATGATGAAAAAGATGAAAGAGCAGCAATGCTTAATAAG 353

33 rgGluPrOvalValAlaPThrPhcGluAlaGlyLysAsnCysAlapro 49

```

alignment\_scores:  
 Quality: 466.50 Length: 129  
 Ratio: 4.092 Gaps: 3  
 Percent Similarity: 88.372 Percent Identity: 70.543

alignment\_block:  
 US-09-327-750D-33 x AAH45143 ..

Align seg 1/1 to: AAH45143 from: 1 to: 792

1 MetGluSerLysAspGlnGlyAlaLysAsn...LeuAsnMetGluAsnAs 16  
 173 ATGGAGTCCCAAGAGAACTAGCTAAACAGCTCTCAGCATGGGAAATGC 222  
 16 pHisGlnLysGluGluLysGluGluLysProGlnAspThrIleLysA 33  
 223 CAACCAAGAAATGACAGAAAGAG.....CAAGTTGCTAATAAAG 263  
 33 rgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49  
 264 GGGAGCCCTTGGCCCTCCT...TTGGATGCTGGTGAATCTGTGCCT 310  
 50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66  
 311 AGAGAAATCGTAGCGGTTCCGCGTTAGCAGGCCCATCTGCAGTATAG 360  
 66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83  
 361 ATGGATATGATGATAGCTTGGAGAACACAGGCAAGGATGAGAGAAG 410  
 83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99  
 411 AGAATATGAAAGGATTGGGAGGGGTGAGACAGCTGATGGAAAGCTG 460  
 100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116  
 461 AGGAAAGAGGTTGAGTCTAGTCTCGCGGCGAGTCAGCAGTACGCCCC 510  
 116 oHisHisAspHisAspGluPheCysLeuMetPro 128  
 511 TCACCATGACCATCATGATGAGTTTGCCTTATGCC 547

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT: AAC03880

seq\_documentation\_block:  
 ID AAC03880 standard; cDNA; 662 BP.  
 XX AAC03880;  
 AC AAC03880;  
 XX 06-OCT-2000 (first entry)  
 XX Human secreted protein 5' EST, SEQ ID NO: 3878.  
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 XX gene therapy; chromosome mapping; ss.  
 KW Homo sapiens.  
 XX EP1033401-A2.  
 PN 06-SEP-2000.  
 PD 21-FEB-2000; 2000EP-0200610.  
 PF 26-FEB-1999; 99US-0122487.  
 PR (GEST ) GENSET.  
 PA Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI WPI: 2000-500381/45.  
 DR P-PSDB; AAG03874.  
 DR XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX Claim 1: SEQ ID 3878; 71pp + CD-ROM; English.  
 XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.  
 XX Sequence 662 BP; 177 A; 144 C; 192 G; 137 T; 12 Other;

alignment\_scores:  
 Quality: 455.00 Length: 129  
 Ratio: 4.027 Gaps: 2  
 Percent Similarity: 87.597 Percent Identity: 66.667

alignment\_block:  
 US-09-327-750D-33 x AAC03880 ..

Align seg 1/1 to: AAC03880 from: 1 to: 662

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16  
 207 ATGGAGTCCCAAGAGAGGAGCAGCGCTTAAACAATCTCATCTGGGAAATGT 256  
 16 pHisGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysA 33  
 257 CAACCAAGAAATGATGAAAGAGATGAAAGGAGCAAGTGCCTAATAAAG 306  
 33 rgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49  
 307 GGGAGCCCTTGGCCCTACCT...TTGATGTYTRGTGAATACTGTGTGCCT 353  
 50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66  
 354 AGAGAAATCGTAGCGGTTCCGCGTTAGCAGGCCCATCTGCAGTATAG 403  
 66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83  
 404 ATGGGATATGATGATGATGCTTGGAGAACCCACAGGCAAGGATGANAGAG 453  
 83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99  
 454 AGAATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAAGCTG 503  
 100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116  
 504 AGGAAAGAGCAGTTGAGTCTAGTCTCGCGGCGAGTCAGCAGTACGCCCC 553  
 116 oHisHisAspHisAspGluPheCysLeuMetPro 128  
 554 TCACCATGACCATCATGATGAGTTTGCNNWATGCC 590

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT: AAC03879  
 seq\_documentation\_block:  
 ID AAC03879 standard; cDNA; 698 BP.  
 XX AAC03879;  
 AC AAC03879;  
 XX

DT 06-OCT-2000 (first entry)  
XX Human secreted protein 5' EST, SEQ ID NO: 3877.  
DE  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
XX  
XX EP1033401-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-0200610.  
XX  
XX 26-FEB-1999; 99US-0122487.  
XX  
XX (GSEST ) GENSET.  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
PI  
XX WPI: 2000-500381/45.  
XX  
XX P-PSDB; AAG03873.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 3877; 71pp + CD-ROM; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
XX Sequence 698 BP; 187 A; 151 C; 210 G; 139 T; 11 other;

alignment\_scores:  
Quality: 455.00 Length: 129  
Ratio: 4.027 Gaps: 2  
Percent Similarity: 87.597 Percent Identity: 66.667

alignment\_block:  
US-09-327-750D-33 x AAC03879 ..  
Align seg 1/1 to: AAC03879 from: 1 to: 698

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16  
243 ATGGAGTCCAAAGAGGACGCGGTTAAACAATCTCATCGTGGAATGCT 292  
16 pHisGlnLysLysGluGluLysGluLysProGlnAspThrIleLysA 33  
293 CAACCGAGAAATGATGAAAGAGATGAAAGAGGACGATGCTGCTAATAAG 342  
33 rgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49  
343 GGGAGCCCTGGCCCTACCT...TTGATGTYTGTGTAATCTGTGCTCT 389  
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66  
390 AGAGGAATCGTAGGCGGTTCCGCGTTAGGACGCCCATCTGCAGTATAG 439

66 gTrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
440 ATGGATATGATGATAGCTTTGGAGACACACAGGCAAGATGANAAG 489  
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
490 AGAATATGAAAGGATTGGGAGGAGGTGAGACAGCTGATGAAAAGCTG 539  
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
540 AGGGAAAAGCAGCTTGATGTCATAGTCTGCGGCGAGTCAGCACTGACCC 589  
116 ohlHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
590 TCACCATGACCATCATGATGAGTGTTCNNWATGCC 626

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI60367

seq\_documentation\_block:

ID: AAI60367 standard; cDNA: 858 BP.

XX AC AAI60367;

XX DT 22-OCT-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 4356.

XX KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI: 2001-442253/47.

XX DR P-PSDB; AAM41211.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

XX PS Claim 1; SEQ ID NO 4356; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX CC the encoded polypeptides (AAM38642-AAM42213) with nontropic.

XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful

XX CC in gene therapy. A composition containing a polypeptide or polynucleotide

XX CC of the invention may be used to treat diseases of the peripheral nervous

XX CC system, such as peripheral nervous injuries, peripheral neuropathy and

XX CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 858 BP: 236 A; 175 C; 241 G; 206 T; 0 other:

alignment_scores:		
Quality:	455.00	Length: 130
Ratio:	3.889	Gaps: 3
Percent Similarity:	90.000	Percent Identity: 67.692

```

alignment_block:
  US-09-327-750D-33 x AAI60367  ..
  Align seg 1/1 to: AAI60367 from: 1 to: 858

1  MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetClnAsnAs 16
215  ATGGAGTCCAAAGAGAACGAGCGTTAAACAATCTCATCGTGGAAAAATGT 264

16  pHisGlnLysLysGluGluLysGlnAspThrIleLysA 33
265  CAACCAGGAAATCATCAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 314

33  rgGluProValValAlaProThrPheClnAlaGlyLysAsnCysAlaPro 49
315  GGGAGACCGTTGGCCCTACCT...TTGAATGTTAGTGAATCTGTGTGCCT 361

50  ArgGlyGlyArgArgPheArgValArgGlnProCileSerHisIyrrAr 66
362  AGAGAAACCGTAGCGCGGTTCCGCGTTAGCAGCCCATCTTCGACGATTAG 411

66  gTrpAspLeuMethHisArgValGlyGluProGlnGlyArgMetArgGluG 83
412  ATGGACATAATGCATAGGCTTGAGAGGCCACAGGCAAGATGAGAGAGG 461

83  luAsnValGlnArgPheGlyClnAspMetArgGlnLeuMetGluLysLeu 99
462  AGAATATGGAAGAGATTGGGGAGGAGGTGAGACAGCTGATGCAAAAAGCTG 511

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
512  AGGGAAAGACAGTTGAGTCATAGTCTCGGGCAGTCAGCACATGATCCCCC 561

116 oHisHisAspHisAspGluPheCys..LeuMetPro 128
562  TCACCATGACCATCAGCATGAGTTTGCCTTTATGCC 599

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAC10889

seq\_documentation\_block:

ID AAC10889 standard; cDNA; 692 BP.

XX

AAC10889;

XX DE 06-00E-30

06-OCT-2000 (first entry)  
DI  
XX

Human secreted protein 5, EST. SEQ ID NO: 14964

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

gene therapy; chromosome mapping; ss.

XX 50

OS Homo sapiens.  
xy

XX  
PN  
EP1033401-A2

XX  
EF1033401-AZ.

```

PD 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST)
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs
XX diagnostic, forensic, gene therapy and chromosome mapping
XX
XX Claim 1; SEQ ID 14964; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs d
XX mRNAs encoding secreted proteins. No ORF has yet been conc
XX identified within the present sequence. The 5' ESTs were p
XX total human RNAs or polyA+ RNAs derived from 30 different
XX sequences usually correspond mainly to the 3' untranslated
XX of the mRNA because they are often obtained from oligo-dT
XX libraries. Such ESTs are not well suited for isolating cDN
XX derived from the 5' ends of mRNAs and even in those cases
XX cDNA sequences have been obtained, the full 5' UTR is rare
XX 5' ESTs are derived from mRNAs with intact 5' ends and can
XX used to obtain full length cDNAs and genomic DNAs. 5' EST
XX in diagnostic, forensic, gene therapy and chromosome mappi
XX They are used to obtain upstream regulatory sequences and
XX expression and secretion vectors.
XX
XX Sequence 692 BP; 184 A; 151 C; 199 G; 146 T; 12 other;

alignment_scores:
    Quality: 444.50      Length: 125
    Ratio: 4.041         Gaps: 1
Percent Similarity: 88.000 Percent Identity: 64.800

alignment_block:
US-09-327-750D-33 x AAC10889 ..

Align seg 1/1 to: AAC10889 from: 1 to: 692

4 LysAspGlnGlyAlaLysAsnLeuAsnMeGluAsnAspHisGlnLysLy 20
249 AGAAGAACGAGCGTTTAAACAATCTCATCGTGCAGAAATGTC AACCCAGGAAA 298

20 sGluGluGlyGluGlyProGlnAspThrIleLysArgGluProValV 37
299 TGATGAAAAGATGAAAAGGAGCAGTGWGCTATATAAGGGGAGCCCTTG 348

37 alAlaProThrPheGluAlaGlyLysAsnCysAlaProArgGlyGlyArg 53
349 CCCTACCT...TTGATGTYTRGTCAACTGTCGTGCTAGAGGAATCGT 395

54 ArgArgPheArgValArgGlnProIleSerHisTyrArgTrpAspLeuMe 70
396 AGGCGGTTCCGCGTTAGGCAGCCCATCTGCAGTATAGATGGGATATGAT 445

70 tHisArgValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnA 87
446 GCATAGGCTTGGAGAACCCAGGCAAGGATGANAGAGAGAATATCGAA 495

87 rgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArgGluArgGln 103
496 GGATTGGGAGGAGTGAGACAGCTGATGAAAGCTGAGGGAAAGACGAG 545

104 LeuSerHisSerLeuArgAlaValSerThrAspProProlHisAspHis 120
546 TTGAGTCATAGTCTCGGGCAGTCAGCAGCTGACCCGCCCTCACCATCACCA 595

```

Sequence 700 BP; 177 A; 188 C; 203 G; 132 T; 0 other;''

assignment_scores:	
Quality:	173.00
Ratio:	2.190









**THIS PAGE BLANK (USPTO)**

## alignment\_block:

```
US-09-327-750D-33 x AAF21748 ..
Align seg 1/1 to: AAF21748 from: 1 to: 917

18 GlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysArgG1 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 CAGGAAACCAAGCAGATGGAGCGCTCATCGACAATGAGAGAGACCG 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 uProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArgG 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 CCCTTTG.....GGAGGAGGTGAAGGCCACCGCCT...G 387
XX

51 LyGlyArgArgArgPheArgValArgGlnProIleSerHisTyrArgTir 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 CAGGAATCGACGGGACAGCGCTCGCGACTTCGCCCTTATTTTCGATGG 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 AspLeuMethHisArg...ValGlyGluProGlnGlyArgMetArgGluG1 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 GCCATACCCCAATAGGCAGATCAATGATGGGATGGGTGA...CATGGAGA 484
XX

83 uAsnValGlnArgPheClyGluAspMetArgGlnLeuMetGluLysLeuA 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
485 TGATATGGAATATTCATCGAGGAGATGAGAGAAATCAGAAGAAACTTA 534
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 rGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp....Pro 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
535 GGAGCTGCGAGTTGAGGAATGTCTGCGTATCCTTATGGGGGAGCTCTCT 584
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 ProHisHisAspHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
585 AATCACCACCATCATCATGATGATTTTGCCTTATGCCT 623
```

seq\_name: /SID32/gcgdata/geneseq/geneseq/NA2000.DAT:AA01005

## seq\_documentation\_block:

```
ID AA01005 standard; cDNA; 532 BP.
XX
AC AA01005;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1003.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG00999.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS Claim 1; SEQ ID 1003; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
```

```
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
```

SQ Sequence 532 BP; 151 A; 121 C; 163 G; 91 T; 6 other;

## alignment\_scores:

```
Quality: 135.50 Length: 120
Ratio: 1.831 Gaps: 5
Percent Similarity: 61.667 Percent Identity: 32.500
```

## alignment\_block:

US-09-327-750D-33 x AAC01005 ..

Align seg 1/1 to: AAC01005 from: 1 to: 532

```
4 LysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHisGlnLysLy 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 AAACAACCCAGAAAAAAATCTCATCTATGGCAAAATATTCACCCAGGAAA 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
20 sGluGluLysGluGluLysProGlnAspThrIleLysArgGluProValV 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 CGAAGAGATGGACGACGCTATGCAGATGAGAGAGAGACCGCCTTTG. 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
37 alaProThrPheGluAlaGlyLysAsnCysAlaProArgGlyGlyArg 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 .....GGAGGAGGTGAAGGCCACCGCCT...GCAGGAAAT 323
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 ArgArgPheArgValArgGlnProIleSerHisTyrArgTrpAspLeuMe 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 CGACGGGAMAGCTCGCGAYTTGCCCTTAATTTTCGATGGGCCATACC 373
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 tHisArg...ValGlyGluProGlnGlyArgMetArgGluGluAsnValG 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 CAATAGGCACATCAATGATGGGATGGT...GRAGATGGAGATGATATGG 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
86 InArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArgGluArg 102
:: ||| |||:|||||:|||||:|||||:|||||:|||||:
421 AAATATTCATGGAGGAGATGAGAGAAATCAGAAGAAACCTTAGGGAGCTG 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 GlnLeuSerHisSerLeuArg...AlaValSerThrAspProProHisH1 118
||| ::| |||||:|||||:|||||:|||||:|||||:
471 CAKTTGAGGAATGTCTGCGTATCTTATGGGGGAKCTCTCTAATACCA 520
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 sAspHisHis 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
521 TGACCATCAT 530
```

OM of: US-09-327-750D-33 to: GenEmbl.\* out\_format : pfs  
Date: Mar 11, 2002 3:34 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet.p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09327750/runat\_11032002.101154.20319/app\_query.fasta\_1.1472  
-DB=GenEmbl -QFWT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09327750\_@CGNL\_1\_8673  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-327-750D-33  
Query length: 128  
Database: GenEmbl.\*  
Database sequences: 1472140  
Database length: 341344837  
Search time (sec): 4557.230000

score_list:	Seq	to	Strd	Orig	zScore	EScore	Len	Documentation
gb_to:AF097437	+	646.00	986.65	1.1e-46	2269	1.1e-46	2269	AF097437 Mus musculus Bex1 prc
gb_to:AF097438	+	640.00	985.02	1.4e-46	835	1.4e-46	835	AF097438 Mus musculus brain exp
gb_to:AF051347	+	640.00	984.99	1.4e-46	838	1.4e-46	838	AF051347 Mus musculus REX-3 mRN
gb_to:AF097439	+	601.50	926.17	2.6e-43	785	1.4e-43	785	AF097439 Mus musculus brain exp
gb_to:BC003254	+	582.00	886.16	4.4e-41	2901	1.4e-41	2901	BC003254 Mus musculus, Similar
gb_pr:AF183416	+	471.50	727.38	3.1e-32	642	1.4e-32	642	AF183416 Homo sapiens ovarian q
gb_pr:AF220189	+	471.50	725.44	3.9e-32	828	1.4e-32	828	AF220189 Homo sapiens uncharact
gb_pr:HS198P4	-	471.50	726.77	1.6e-30	35714	1.6e-30	35714	AL008708 Human DNA sequence f
gb_pr:AF251053	+	468.00	720.87	7.1e-32	744	1.4e-32	744	AF251053 Homo sapiens X-linked
gb_pr:AX078272	+	468.00	719.43	8.5e-32	898	1.4e-32	898	AX078272 Sequence 76 from Paten
gb_pr:HSV870H8	+	468.00	692.37	2.7e-30	31321	1.4e-30	31321	AF183416 Human DNA sequence f
gb_pr:AL133348	-	468.00	690.40	3.5e-30	40584	1.4e-30	40584	AL133348 Human DNA sequence f
gb_htg:HSU80B1	+	468.00	690.32	3.6e-30	41029	1.4e-30	41029	AL022169 Homo sapiens Chromos
gb_pr:AF237783	+	466.50	718.09	1.0e-31	791	1.4e-31	791	AF237783 Homo sapiens brain-exp
gb_sts:G24641	-	311.00	481.91	1.4e-18	504	1.4e-18	504	G24641 human STS WI-11354, sequ
gb_pr:HSV351F8	+	195.50	269.60	9.7e-07	45678	1.4e-07	45678	AF183416 Human DNA sequence f
gb_pr:HUM00C	+	187.50	287.27	1.0e-07	891	1.4e-07	891	M38188 Human unknown protein fr
gb_pr:AF187064	+	187.50	287.27	1.0e-07	891	1.4e-07	891	AF187064 Homo sapiens p75NTR-as
gb_pr:BC003190	+	186.50	286.62	1.1e-07	793	1.4e-07	793	BC003190 Homo sapiens, p75NTR-a
gb_to:AF187066	+	174.00	268.30	1.1e-06	700	1.4e-06	700	AF187066 Mus musculus p75NTR-as
gb_to:AF097440	+	174.00	266.79	1.4e-06	854	1.4e-06	854	AF097440 Mus musculus brain exp
gb_sts:G35294	-	170.00	265.06	1.7e-06	477	1.7e-06	477	G35294 human STS SHGC-37409, se
gb_pr:AK000959	+	170.00	257.85	4.4e-06	1229	1.4e-06	1229	AK000959 Homo sapiens cDNA FLJ
gb_pr:AX100231	+	170.00	257.06	4.8e-06	1364	1.4e-06	1364	AX100231 Sequence 32 from Paten
gb_pr:HS635G19	+	170.00	227.10	0.0002	69648	1.4e-02	69648	AL035494 Human DNA sequence f
gb_to:AF187065	+	148.00	230.52	0.0001	519	1.4e-01	519	AF187065 Rattus norvegicus p75N
gb_sts:G72708	-	142.50	223.64	0.0004	421	1.4e-01	421	G72708 MARC 4953-4954:991939031
gb_da:SC61A	+	102.00	127.34	81.34	36028	1.4e-01	36028	AL442120 Streptomyces coelicol
gb_sts:G23364	+	100.50	159.86	1.25	372	1.4e-01	372	G23364 human STS WI-15922, sequ
gb_pr:AB040971	+	97.50	136.41	25.41	4408	1.4e-01	4408	AB040971 Homo sapiens mRNA for
gb_htg:AC021617	+	94.50	103.23	1.8e+03	187075	1.4e+03	187075	AC021617 Homo sapiens clone
gb_pat:AC25464	+	93.50	142.84	11.14	844	1.4e+03	844	A25464 B.pertussis STX pilin ge
gb_da:BFIMX	+	93.50	142.83	11.15	845	1.4e+03	845	Y00556 Bordetella pertussis fil
gb_da:BFIMX	+	93.50	139.46	17.18	1315	1.4e+03	1315	X74118 B.bronchiseptica paten
gb_pat:AF73843	+	93.50	139.46	17.18	1315	1.4e+03	1315	AF73843 Sequence 1 from Paten
gb_pat:AR029509	+	93.00	130.97	51.07	3624	1.4e+03	3624	AR029509 Sequence 6 from paten
gb_pat:AR098462	+	93.00	130.97	51.07	3624	1.4e+03	3624	AR098462 Sequence 6 from paten
gb_pat:141422	+	93.00	130.97	51.07	3624	1.4e+03	3624	I41422 Sequence 6 from paten
gb_pr:MACP53A	+	92.00	133.28	37.94	2184	1.4e+03	2184	L20442 Rhesus monkey p53 mRNA,
gb_da:HVU95372	-	92.00	122.97	142.38	8455	1.4e+03	8455	U95372 Haloferax volcanii plas

gb\_ba:AF007125 + 91.00 115.68 362.86 17992 ! AF007125 Mycobacterium tub  
gb\_ba:MTU00024 - 91.00 111.49 620.87 31175 ! U00024 Mycobacterium tuber  
gb\_ba:MSGY2 - 91.00 111.49 620.89 31176 ! A000009 Mycobacterium tub  
gb\_ba:WTCY349 + 91.00 108.95 860.19 43523 ! Z83018 Mycobacterium tuber  
gb\_htg:AP001332 + 91.00 100.23 2.6e+03 136678 ! AP001332 Homo sapiens chr  
seq\_name: gb\_to:AF097437  
seq\_documentation\_block:  
LOCUS AF097437 2269 bp DNA ROD 05-AUG-1999  
DEFINITION Mus musculus Bex1 protein (Bex1) gene, complete cds.  
ACCESSION AF097437  
VERSION AF097437.1 GI:5702155  
KEYWORDS house mouse.  
SOURCE  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2269)  
AUTHORS Brown,A.L. and Kay,G.F.  
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,  
is a member of a novel gene family on the mouse X chromosome  
Hum. Mol. Genet. 8 (4), 611-619 (1999)  
JOURNAL 99172070  
MEDLINE 10072429  
PUBMED  
REFERENCE 2 (bases 1 to 2269)  
AUTHORS Brown,A.L. and Kay,G.F.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-1998) Cancer Unit, Queensland Institute of  
Medical Research, Herston Rd, Brisbane, Qld 4030, Australia  
FEATURES  
source  
1..2269  
/organism="Mus musculus"  
/strain="129SVJ"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/map="near Pip"  
/tissue\_type="Liver"  
join(409..509,866..942,1263..1873)  
/gene="Bex1"  
/product="Bex1 protein"  
409..1873  
/gene="Bex1"  
/note="expressed in brain; X-linked"  
1268..1654  
/gene="Bex1"  
/codon\_start=1  
/product="Bex1 protein"  
/protein\_id="AA047168.1"  
/db\_xref="GI:5702156"  
translation="MESKDGKVLNNMNDHKKKEEKPODTRREPAVALTSEAG  
KNCAPGGRRRFRVROPPIAHYRDLQVRGEPQGRMRREENVQRGGDVRLMEKLRR  
QLSHSLRAVSTDPHDDHDEFLMP"  
BASE COUNT 563 a 554 c 705 g 447 t  
ORIGIN  
alignment\_scores:  
Quality: 646.00 Length: 128  
Ratio: 5.168 Gaps: 0  
Percent Similarity: 97.656 Percent Identity: 92.969  
alignment\_block:  
US-09-327-750D-33 x AF097437 ..  
Align seg 1/1 to: AF097437 from: 1 to: 2269  
1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspH1 17  
|||||  
1268 ATGGAGTCCAAAGATCAACGGCTGAAATCTCAACATCGAATGACCA 1317  
17 sGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysArgG 34  
|||||

333

```

cnc
220 570
/cell_line="F9 teratocarcinoma"

```

/note="retinoic acid reduced gene expression"

/codon\_start=1  
/product="REX-3"  
/protein\_id="AAC61929.1"  
/db\_xref="GI:3510643"  
/translation="MENDHQKKEEKEKPDQIRREPAVALISEAGKNCAPRGGRRRF  
RYPQIAHYRMDLMQVRGEPQGRMEENVQRFQDLMEKLRERQLSHSLRAVSTD  
PPHDDHDFCLMP"  
BASE COUNT 238 a 181 c 251 g 168 t  
ORIGIN

alignment\_scores:  
Quality: 640.00 Length: 128  
Ratio: 5.161 Gaps: 0  
Percent Similarity: 96.875 Percent Identity: 92.188

alignment\_block:

US-09-327-750D-33 x AF051347

Align seg 1/1 to: AF051347 from: 1 to: 838

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHi 17  
193 ATGGAGTCCAAAGATCAAGCGTGAATAATCTCAACATGGAGAATGACCA 242  
17 sGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysArg 34  
243 TCAGAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 292  
34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50  
293 AGCCAGCTGTGGCCCTCATCTCCGAGCTGGCAAAACTGCGCCCTAGA 342  
51 GlyGlyArgArgArgPheArgValArgGlnProIleSerHisTyrArgTr 67  
343 GGAGGTCCGACGGCGTTTCGGGGTTCGGCAGCCCATCGCTCACTATAGT 392  
67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluAla 84  
393 GGACCTCATCAGAGGGTTGGGAGCCCGCAGGAGGAGGAGGAGGAGGAG 442  
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100  
443 ACGTACAGAGGTTTGGGGGTGATGTGACAGCTCATGGAGAGCTCAGG 492  
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProHi 117  
493 GAAAGGCAGCTGAGCCACAGCCTCGGGCGGTTAGCAGTACCCGCGCTCA 542  
117 sHisAspHisAspGluPheCysLeuMetPto 128  
543 TCATGACCACCATGATGAGTTTGCCTCATGCCCC 576

seq\_name: gb\_ro:AF097439

seq\_documentation\_block:  
LOCUS AF097439 785 bp mRNA ROD 13-APR-1999  
DEFINITION Mus musculus brain expressed x-linked protein 2 (Bex2) mRNA,  
complete cds.  
ACCESSION AF097439  
VERSION AF097439  
KEYWORDS AF097439.1 GI:4580591  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 785)  
AUTHORS Brown,A.L. and Kay,G.F.  
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,  
is a member of a novel gene family on the mouse X chromosome  
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)  
MEDLINE 99172070

REMARK

Erratum: [[published erratum appears in Hum Mol Genet 1999  
May;8(5):943]]  
2 (bases 1 to 785)  
AUTHORS Brown,A.L. and Kay,G.F.  
TITLE Direct Submission  
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of  
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia  
FEATURES  
Location/Qualifiers  
1..785  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="dbEST:W48832"  
/chromosome="X"  
/map="near plp"  
/dev\_stage="embryo; 15.5 dpc"  
1..785  
/gene="Bex2"  
139..528  
/gene="Bex2"  
/codon\_start=1  
/product="brain expressed x-linked protein 2"  
/protein\_id="AAD24430.1"  
/db\_xref="GI:4580592"  
/translation="MESKVEQGVKNLNMENDHQEKEEKEKPODASKRDPVIALPFEA  
GDYVYRGRRRRFVRQPIVHYRMDLMHRVGEPOGRMEENVQRFQDLMEKLRERQLMEKLR  
ROLSHSLRAVSTDPPHDDHDFCLMP"  
BASE COUNT 212 a 163 c 235 g 175 t  
ORIGIN

alignment\_scores:

Quality: 601.50 Length: 129  
Ratio: 4.971 Gaps: 1  
Percent Similarity: 93.798 Percent Identity: 86.822

alignment\_block:

US-09-327-750D-33 x AF097439

Align seg 1/1 to: AF097439 from: 1 to: 785

1 MetGluSerLys...AspGlnGlyAlaLysAsnLeuAsnMetGluAsnAs 16  
139 ATGGAGTCCAAAGTGGAAACAGCGTCAAAATCTCAACATGGAGAATGA 188  
16 pHisGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysA 33  
189 CCATCAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 238  
33 rgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49  
239 GGGATCCGATTGTGGCCCTGCCTTTCGAAGCTGGAGACTACTACGTCCT 288  
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66  
289 AGAGGAGTCCGACGGCGTTCCGGGTTCCGAGCCCATCTGTCACACTACAG 338  
66 gTrpAspLeuMethisArgValGlyGluProGlnGlyArgMetArgGluG 83  
339 ATGGGACCTGATGATAGGGTTGGGAGCCCGCCAGGAGGATGAGAGAGG 388  
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99  
389 AGAACGTACAGAGGTTGGGGATGATGTGACACAGCTCATGGAGAACTG 438  
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116  
439 AGGAAAGGAGGAGTGGAGCCACAGCTCGGGCGGTAGCACTGACCCGCC 488  
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128  
489 TCATCATGACCACCATGATGAGTTTGCCTTATGCC 525





/translation="MESKEKRAVNSLSMENANOBNEKEQVANKGEPLALPLDAGEYC  
VPGNRRFRVRPILQYRWDMMHRLGEPQARMREENMERIGEVRQMLEKREKQLS  
HSLRAVSTDPHHDHDFCLMP"  
BASE COUNT 204 a 118 c 157 g 163 t  
ORIGIN

alignment\_scores:  
Quality: 471.50 Length: 129  
Ratio: 4.100 Gaps: 3  
Percent Similarity: 89.147 Percent Identity: 70.543

alignment\_block:  
US-09-327-750D-33 x AF183416 ..

Align seg 1/1 to: AF183416 from: 1 to: 642

```
1 MetGluSerLysAspGlnGlyAlaLysAsn...LeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 ATGGAGTCCAAAGAGAAACGACGAGTAAACAGTCTCAGCATGCAAAATGC 52
16 pHisGlnLysLysGluGluGluGluLysProGlnAspThrIleLysA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
53 CAACCAAGAAATGAAGAAAGAG.....CAAGTTGCTTAATAAAG 93
33 rgGluProValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
94 GGGAGCCCTGGCCCTCCCT...TTGGATGCTGGTGAATACTGTGTCCT 140
50 ArgGlyClyArgArgPheArgValArgGlnProIleSerHisTyra 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
141 AGAGGAATCGTAGGCGGTTCGCGTGTAGCGAGCCCATCTGCAGTATAG 190
66 gTrpAspLeuMetHisArgValGlyGluAspMetArgGlnLysLeu 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
191 ATGGATATGATCATAGGCTTGGAGAACACAGGCAAGATGAGAGAAG 240
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
241 AGAATATGGAAGAGATTGGGAGAGGTGAGACAGCTGATGGAAGAGCTG 290
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 AGGGAAGACAGTTGATGTCATAGTCTCGCGGCAGTCAGCACTGACCC 340
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 TCACCATGACCATCATGATGAGTTTGGCTTATGCCCC 377
```

seq\_name: gb\_pr:AF220189

seq\_documentation\_block:  
LOCUS AF220189 828 bp mRNA PRI 04-MAY-2000  
DEFINITION Homo sapiens uncharacterized hypothalamus protein HBEX2 mRNA,  
complete cds.  
ACCESSION AF220189  
VERSION AF220189.1 GI:7689028  
KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.

TITLE A novel gene expressed in human hypothalamus

JOURNAL Unpublished

REFERENCE

AUTHORS Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.

TITLE Direct Submission

JOURNAL Submitted (30-DEC-1999) Chinese National Human Genome Center at

Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,

Shanghai 201203, P. R. China

Location/Qualifiers

FEATURES

source

1..828

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/tissue\_type="hypothalamus"

167..544

/codon\_start=1

/product="uncharacterized hypothalamus protein HBEX2"

/protein\_id="AAF67654.1"

/db\_xref="GI:7689029"

/translation="MESKEKRAVNSLSMENANOBNEKEQVANKGEPLALPLDAGEYC

VPGNRRFRVRPILQYRWDMMHRLGEPQARMREENMERIGEVRQMLEKREKQLS

HSLRAVSTDPHHDHDFCLMP

BASE COUNT 256 a 170 c 221 g 181 t

ORIGIN

alignment\_scores:

Quality: 471.50 Length: 129

Ratio: 4.100 Gaps: 3

Percent Similarity: 89.147 Percent Identity: 70.543

alignment\_block:

US-09-327-750D-33 x AF220189 ..

Align seg 1/1 to: AF220189 from: 1 to: 828

```
1 MetGluSerLysAspGlnGlyAlaLysAsn...LeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 ATGGAGTCCAAAGAGAAACGACGAGTAAACAGTCTCAGCATGCAAAATGC 216
16 pHisGlnLysLysGluGluGluGluLysProGlnAspThrIleLysA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 CAACCAAGAAATGAAGAAAGAG.....CAAGTTGCTTAATAAAG 257
33 rgGluProValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
258 GGGAGCCCTGGCCCTCCCT...TTGGATGCTGGTGAATACTGTGTCCT 304
50 ArgGlyClyArgArgPheArgValArgGlnProIleSerHisTyra 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
305 AGAGGAATCGTAGGCGGTTCGCGTGTAGCGAGCCCATCTGCAGTATAG 354
66 gTrpAspLeuMetHisArgValGlyGluAspMetArgGlnLysLeu 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
355 ATGGGATATGATCATAGGCTTGGAGAACACAGGCAAGATGAGAGAAG 404
83 luAsnValGlnArgPheGlyGluAspMetArgGlnLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
405 AGAATATGGAAGAGATTGGGAGAGGTGAGACAGCTGATGGAAGAGCTG 454
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
455 AGGGAAGACAGTTGATGTCATAGTCTCGCGGCAGTCAGCACTGACCC 504
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
505 TCACCATGACCATCATGATGAGTTTGGCTTATGCCCC 541
```

seq\_name: gb\_pr:HS198P4

seq\_documentation\_block:

LOCUS HS198P4 35714 bp DNA PRI 16-DEC-2000

DEFINITION Human DNA sequence from clone RPI-198P4 on chromosome Xq22 Contains

a CpG Island, complete sequence.

ACCESSION AL008708

VERSION AL008708.4 GI:4826445

KEYWORDS HTG; CpG Island.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 35714)

AUTHORS Corby,N.

TITLE  
JOURNAL

Direct Submission  
Submitted (21-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerquest@sanger.ac.uk

## COMMENT

On May 13, 1999 this sequence version replaced gi:4582117.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.

This sequence was generated from part of bacterial clone contigs of  
human chromosome X, constructed by the Sanger Centre Chromosome X  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/ChrX>

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RPI-198P4 is  
from the library RPCI-1 constructed at the Roswell Park Cancer  
Institute by the group of Pieter de Jong. For further details see  
<http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone  
RPI-198P4. It may be shorter because we only sequence overlapping  
sections once, or longer because we arrange for a small overlap  
between neighbouring submissions.  
The true left end of clone LLOXNC01-221P2 is at 35615 in this  
sequence. The true right end of clone LLOXNC01-73E8 is at 100 in  
this sequence.

## FEATURES

## source

Location/Qualifiers  
1..35714

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"

/map="q22"

/clone="RPI-198P4"

/clone.lib="RPCI-1"

1..258

/note="L1P3 repeat: matches 5666. .5925 of consensus"  
256. .537

/note="L1P3 repeat: matches 5391. .5672 of consensus"  
1491. .1749

/note="AluSc repeat: matches 39. .297 of consensus"  
2182. 3810

/note="L1P3 repeat: matches 4533. .6150 of consensus"  
5169. .6687

/note="L1MA2 repeat: matches 4798. .6308 of consensus"  
6976. .7665

/note="L1MA2 repeat: matches 4113. .4798 of consensus"  
8039. 9213

/note="L1MA2 repeat: matches 2945. .4113 of consensus"  
9523. .10437

/note="L1MA2 repeat: matches 2016. .2945 of consensus"  
10736. .11494

/note="L1MA2 repeat: matches 1262. .2016 of consensus"  
11807. .12251

/note="L1MA2 repeat: matches 310. .755 of consensus"  
12259. .12331

/note="HERVH21 repeat: matches 5. .77 of consensus"  
12332. .12796

/note="L1TR21A repeat: matches 1. .490 of consensus"  
16162. 16371

/note="105 copies 2 mer tt 55% conserved"  
17121. .17424

/note="MER33 repeat: matches 1. .322 of consensus"

repeat\_region

18259. .18290  
/note="16 copies 2 mer ca 87% conserved"

repeat\_region

18505. .18672  
/note="MER63A repeat: matches 1. .181 of consensus"  
19478. .19814

repeat\_region

/note="MER2 repeat: matches 1. .345 of consensus"  
21061. .21101

repeat\_region

/note="HERVH21 repeat: matches 42. .85 of consensus"  
21247. .21781

misc\_feature

/note="MLT1E repeat: matches 44. .568 of consensus"  
24465. .24492

misc\_feature

/note="Single clone region"  
Single clone region"

misc\_feature

24531. .24562  
/note="Weak data"

misc\_feature

24563

/note="Single clone region"  
Single clone region"

misc\_feature

24711. .24865  
/note="region covered by subclones from a PAC DNA PCR

product only at 12x coverage.  
region covered by subclones from a PAC DNA PCR product only

at 12x coverage."

misc\_feature

24758. .25460  
/note="CpG island"

repeat\_region

/evidence="not\_experimental"  
24875. .24940

repeat\_region

/note="33 copies 2 mer cc 66% conserved"  
26795. .26926

repeat\_region

/note="66 copies 2 mer ct 63% conserved"  
27088. .27133

repeat\_region

/note="23 copies 2 mer tg 100% conserved"  
27461. .27777

repeat\_region

/note="L1MB8 repeat: matches 5851. .6171 of consensus"  
27802. .28232

repeat\_region

/note="L1MC/D repeat: matches 5192. .5403 of consensus"  
28536. .28930

repeat\_region

/note="L1MC/D repeat: matches 4802. .5192 of consensus"  
29240. .29341

repeat\_region

/note="TIGGER1 repeat: matches 484. .587 of consensus"  
29603. .30740

repeat\_region

/note="TIGGER1 repeat: matches 587. .1757 of consensus"  
31027. .31365

repeat\_region

/note="TIGGER1 repeat: matches 1757. .2105 of consensus"  
31664. .31691

repeat\_region

/note="TIGGER1 repeat: matches 2105. .2132 of consensus"  
31809. .32015

repeat\_region

/note="TIGGER1 repeat: matches 2132. .2246 of consensus"  
32323. .32389

repeat\_region

/note="TIGGER1 repeat: matches 2246. .2313 of consensus"  
34353. .34636

repeat\_region

/note="MLT1-INTERNAL repeat: matches 607. .889 of  
consensus"

repeat\_region

35149. .35376

repeat\_region

/note="MLT1-INTERNAL repeat: matches 1001. .1233 of  
consensus"

repeat\_region

35233. .35670  
/note="MLT1C repeat: matches 8. .159 of consensus"

BASE COUNT 9708 a 7094 c 6805 g 12107 t

ORIGIN

alignment\_scores:

Quality: 471.50

Ratio: 4.100

Percent Similarity: 89.147

Percent Identity: 70.543

alignment\_block:

US-09-327-750D-33 x HS198P4/rev ..

Align seg 1/1 to reverse of: HS198P4 from: 1 to: 35714

Length: 129

Gaps: 3

```

1 MetGluSerLysAspGlnGlyAlaLysAsn...LeuAsnMetGluAsnAs 16
24358 ATGGAGTCCAAAGAGAACAGCAGCTAAACAGCTCTCAGCATGGAAAATGC 24309
16 pHisGlnLysLysGluGluLysGluLysProGlnAspThrIleLysA 33
24308 CAACCAAGAAATGAAGAAAGGAG.....CAAGTTGCTAATAAAG 24268
33 rgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
24267 GGGAGCCCTTGGCCCTCCCT...TTGGATGCTGGTCACTACTGTGTGCCT 24221
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 66
24220 AGAGGAATCGTAGGCGGTTCGCGCTTAGGCAGCCCATCTCGAGTATAG 24171
66 gtrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
24170 ATGGGATATGATGATGAGCTTGGAGAACACACAGGCAAGGATGAGAGAAG 24121
83 LuAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
24120 AGAATATGAAGAGATTGGGAGGAGTGAGACAGCTGATGGAAAGCTG 24071
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
24070 AGGGAAAGCAGTTGAGTTCATAGTCTGCGGCGAGTCAGCACTGACCCCCC 24021
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
24020 TCACCATGACCATCATGATGATGAGTTTGCCTTATGCCCC 23984
seq_name: gb_pr:AF251053

```

```

seq_documentation_block: 744 bp mRNA PRI 15-APR-2001
LOCUS AF251053
DEFINITION Homo sapiens X-linked protein mRNA, complete cds.
ACCESSION AF251053
VERSION AF251053.1 GI:13625167
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Mao, Y., Xie, Y., Zhou, Z., Zhao, W., Zhao, S., Wang, W., Huang, Y.,
Wang, S., Tang, R., Chen, X. and Wu, C.
TITLE
Direct Submission
JOURNAL
Submitted (29-MAR-2000) Institute of Genetics, School of Life
Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R.
China
FEATURES
Location/Qualifiers
source
1..744
/organism="Homo sapiens"
/db_xref="taxon:9606"
101..487
CDS
101..487
/codon_start=1
/product="X-linked protein"
/protein_id="AAK34943.1"
/db_xref="GI:13625168"
/translation="MESKEERALNNLIVENNQENDEKDEQVANKGEPLALPLNYS
EYCVPRGNRRFRVROPILQYRWIDIMHRLGEPQARMREENRIGEEVRLMEKLREK
QLSHSLRAVSTDPPHHDHDFCLMP"
BASE COUNT 222 a 134 c 208 g 180 t
ORIGIN
alignment_scores:
Quality: 468.00 Length: 129
Ratio: 4.000 Gaps: 2
Percent Similarity: 90.698 Percent Identity: 68.217
alignment_block:
US-09-327-750D-33 x AX078272
Align seg 1/1 to: AX078272 from: 1 to: 898

```

```

US-09-327-750D-33 x AF251053
Align seg 1/1 to: AF251053 from: 1 to: 744
1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16
101 ATGGAGTCCAAAGAGAACAGCAGCGTTAAACAATCTCATCGTGGAAAATGT 150
16 pHisGlnLysLysGluGluLysGluLysProGlnAspThrIleLysA 33
151 CAACCAAGAAATGAAGAAAGATGAAAAGGCAAGGCAAGTTGCTAATAAAG 200
33 rgGluProValValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
201 GGGAGCCCTTGGCCCTTACCT...TTGAATGTTAGTAATACTGTGTGCCT 247
50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyrAr 56
248 AGAGAAACCGTAGGCGGTTCGCGCTTAGGCAGCCCATCTCGAGTATAG 297
66 gtrpAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluG 83
298 ATGGACATGAATGCATAGCTTGGAGAGCCACAGGCAAGGATGAGAGAG 347
83 LuAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeu 99
348 AGAATATGAAGAGATTGGGAGGAGGTGAGACAGCTGATGAAAAGCTG 397
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
398 AGGGAAAGCAGTTGAGTTCATAGTCTGCGGCGAGTCAGCACTGATCCCC 447
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
448 TCACCATGACCATCATGAGTTTGCCTTATGCCCC 484
seq_name: gb_pat:AX078272
seq_documentation_block: 898 bp DNA PAT 22-FEB-2001
LOCUS AX078272
DEFINITION Sequence 76 from Patent WO0107471.
ACCESSION AX078272
VERSION AX078272.1 GI:13157963
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Hilman, J.L., Lal, P., Tang, Y.T., Yue, H., Au-Young, J., Bandman, O.,
Azinval, Y., Yang, J., Lu, D.A., Baughn, M.R., Patterson, C. and Shah, P.
TITLE
Cell cycle and proliferation proteins
JOURNAL
Patent: WO 0107471-A 76 01-FEB-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source
1..898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 5664154CB1"
BASE COUNT 250 a 186 c 251 g 211 t
ORIGIN
alignment_scores:
Quality: 468.00 Length: 129
Ratio: 4.000 Gaps: 2
Percent Similarity: 90.698 Percent Identity: 68.217
alignment_block:
US-09-327-750D-33 x AX078272
Align seg 1/1 to: AX078272 from: 1 to: 898

```

```

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ATGCAGTCCAAAGGAAGCAGCGTTAAACAATCTCATCTGTTGAAAATGT 303

16 pHisGlnLysGluGlyGluGlyGluGlyProGlnAspThrIleLysA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 CAACGAGAAATGATGAAAAGATGAAAAGGAGCAAGTTGCTAATAAAG 353

33 rgGluProValAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 GGGAGCCCTGGCCCTACCT...TTGATGTTAGTGAATACTGTGTCCT 400

50 ArgGlyArgArgArgPheArgValArgGlnProIleSerHisTyAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 AGAGAAACCGTAGCGGTTCCGCGTTAGCAGCCCATCTGCAGTATAG 450

66 gTriPAspLeuMethHisArgValGlyGluProGlnGlyArgMetArgGlu 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
451 ATGGGACATAATGCATAGCGTTGGAGAGCCACAGGCAAGGATGAGAGG 500

83 luAsnValGlnArgPheGlyCysAspMetArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 AGATATGGAAGAGATTGGGAGGAGGTGAGACAGCTGATGGAAAGCTG 550

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
551 AGGCAAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCACTGATCCCC 600

116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 TCACCATGACCATCAGCATGAGTTTGCCTTATGCCC 637

```

seq\_name: gb\_pr:HSV870H8

```

seq_documentation_block: 31321 bp DNA PRI 23-NOV-1999
LOCUS HSV870H8
DEFINITION Human DNA sequence from cosmid V870H8, between markers DXS366 and
DXS87 on chromosome X contains ESTs.
ACCESSION 270233
VERSION 270233.1 GI:1235542
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 31321)
AUTHORS Whiteley,M.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-1995) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk
COMMENT IMPORTANT: This sequence is not the entire insert of clone V870H8.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone V870H8 is at 1 in this sequence. The
true left end of clone V693A8 is at 31221.
V870H8 is from the human chromosome X-specific cosmid library.
FEATURES
source
location/Qualifiers
1..31321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="GHC-870H8"
/clone_lib="SCCV"
repeat_region 475..701
/feature="L1 element fragment"
repeat_region 802..988
/feature="L1 element fragment"
misc_feature 2533..4138
/feature="match: multiple ESTs"
repeat_region 4189..4246
/feature="29 copies of 2 mer 91 & conserved"

```

```

repeat_region 7496..7874
/feature="L1 element fragment"
repeat_region 8022..8108
/feature="L1 element fragment"
repeat_region 8302..8493
/feature="L1 element fragment"
repeat_region 8585..8671
/feature="L1 element fragment"
repeat_region 8769..9010
/feature="L1 element fragment"
repeat_region 9259..9384
/feature="L1 element fragment"
repeat_region 9624..9731
/feature="MLTIC element fragment"
repeat_region 9781..9948
/feature="MLTIC element fragment"
repeat_region 9890..9948
/feature="MLTIB element fragment"
repeat_region 10418..10592
/feature="Alu repeat: matches 298..108 of consensus"
repeat_region 10593..10688
/feature="Alu repeat: matches 96..1 of consensus"
repeat_region 10698..10781
/feature="L1 element fragment"
repeat_region 11112..11403
/feature="Alu repeat: matches 1..308 of consensus"
repeat_region 11922..12004
/feature="L1 element fragment"
repeat_region 12062..12308
/feature="L1 element fragment"
repeat_region 12423..12707
/feature="Alu repeat: matches 302..1 of consensus"
repeat_region 12710..12916
/feature="L1 element fragment"
repeat_region 12979..13081
/feature="L1 element fragment"
repeat_region 13473..13532
/feature="MLT2A1 element fragment"
repeat_region 15871..15928
/feature="L1 element fragment"
repeat_region 18180..18417
/feature="L1 element fragment"
repeat_region 18778..19056
/feature="L1 element fragment"
repeat_region 22279..22350
/feature="Alu repeat: matches 1..308 of consensus"
repeat_region 22449..22995
/feature="L1 element fragment"
repeat_region 23239..23374
/feature="L1 element fragment"
repeat_region 23371..23606
/feature="MLTIB element fragment"
repeat_region 23395..23541
/feature="MLTIA element fragment"
repeat_region 23705..23778
/feature="MLTIA element fragment"
repeat_region 23805..23960
/feature="L1 element fragment"
repeat_region 24011..24541
/feature="L1 element fragment"
repeat_region 24638..24770
/feature="L1 element fragment"
repeat_region 24829..30975
/feature="Alu repeat: matches 150..1 of consensus"
repeat_region 9462 a 6147 c 6129 g 9583 t
BASE COUNT
ORIGIN
alignment_scores:

```



```

/notes="L2 repeat: matches 1599. .2175 of consensus"
18706. .18755
/notes="L1MB3 repeat: matches 6034. .6083 of consensus"
18756. .19236
/notes="MER6B repeat: matches 1. .486 of consensus"
19237. .21108
/notes="MER66-internal repeat: matches 4808. .6676 of
consensus"
21109. .21163
/notes="MT2B repeat: matches 391. .445 of consensus"
21164. .21555
/notes="MER66B repeat: matches 3. .391 of consensus"
21557. .21662
/notes="L1MB3 repeat: matches 6080. .6185 of consensus"
21665. .21761
/notes="L1M repeat: matches 5299. .5392 of consensus"
21762. .21929
/notes="L1MB8 repeat: matches 5902. .6063 of consensus"
21930. .22212
/notes="AluSg repeat: matches 3. .284 of consensus"
22213. .22321
/notes="L1MB8 repeat: matches 6063. .6173 of consensus"
22333. .22740
/notes="L1MD2 repeat: matches 5867. .6288 of consensus"
22744. .23094
/notes="L1MD repeat: matches 2. .342 of consensus"
23218. .23524
/notes="AluSg repeat: matches 1. .305 of consensus"
23603. .23947
/notes="L1M4 repeat: matches 4916. .5184 of consensus"
23948. .24218
/notes="AluJo repeat: matches 1. .282 of consensus"
24219. .24510
/notes="L1M4 repeat: matches 4601. .4916 of consensus"
24508. .24621
/notes="L1M4 repeat: matches 4374. .4488 of consensus"
24650. .25077
/notes="L1M4 repeat: matches 1. .466 of consensus"
25078. .25914
/notes="L1M4 repeat: matches 3497. .4346 of consensus"
25920. .26741
/notes="L1M4 repeat: matches 2202. .3065 of consensus"
26762. .26906
/notes="L1P7 repeat: matches 6001. .6145 of consensus"
26907. .27135
/notes="L1P5 repeat: matches 5674. .5902 of consensus"
27266. .27509
/notes="L1M4 repeat: matches 2164. .2080 of consensus"
27557. .27769
/notes="L1MEC repeat: matches 253. .460 of consensus"
28120. .28167
/notes="L2 repeat: matches 2703. .2750 of consensus"
29043. .29349
/notes="MER58B repeat: matches 1. .341 of consensus"
30389. .30452
/notes="16 copies 4 mer tata 89% conserved"
complement(30502)
/gene="dj79p11.1"
complement(join(30502. .31137,31514. .31589,32016. .>32134))
/notes="match: ESTs: Em:AA317587 Em:AA358632 Em:A1214048
Em:A1929470 Em:AA319216 Em:AA317625 Em:D81392 Em:AA317418
Em:AA366528 Em:C15186 Em:A1291270 Em:AA160660 Em:AA612605
Em:AW160832 Em:AA384405 Em:AA160657 Em:AA081744
Em:A1291126 Em:C15591 Em:AA384268 Em:A1929703 Em:A1302026
Em:W19547 Em:A1141727 Em:AA946933 Em:A1141583 Em:H61106
Em:W60581 Em:A1826617 Em:AA706545 Em:AA464771 Em:AA862327
Em:W68380 Em:AA641652 Em:AA384269 Em:AA527696 Em:W48778
Em:AA633989 Em:A1659486 Em:AW167301 Em:A1929428
Em:AW379411 Em:C15948 Em:A1745669 Em:AW024587 Em:AW379207;
match: proteins: Tr:Q9Y516 Tr:Q9Y517"
/evidence="not_experimental"
/product="dj79p11.1 (novel protein similar to mouse Bex2

```

```

(gene="brain-expressed X-linked protein 2))"
complement(30502. .32134)
/gene="dj79p11.1"
30507. .30988
/notes="match: STS: Em:G24641; match: STS: Em:T86927"
30508. .30852
/notes="match: STS: Em:R43117; match: STS: Em:G23964"
complement(30518. .30523)
/gene="dj79p11.1"
30541. .30922
/notes="match: STS: Em:H61107"
31543. .32262
/notes="CpG island"
/evidence="not_experimental"
31647. .31730
/notes="42 copies 2 mer cc 64% conserved"
complement(32655. .33316)
/notes="match: GSS: Em:AQ489303"
33027. .34514
/notes="L1ME2 repeat: matches 4631. .6155 of consensus"
34528. .34786
/notes="L1ME3A repeat: matches 4404. .4664 of consensus"
34787. .35085
/notes="AluSg repeat: matches 1. .300 of consensus"
35086. .35412

alignment_scores:
Quality: 468.00 Length: 129
Ratio: 4.000 Gaps: 2
Percent Similarity: 90.698 Percent Identity: 68.217

alignment_block:
US-09-327-750D-33 x AL133348/rev ..

Align seg 1/1 to reverse of: AL133348 from: 1 to: 40584

1 MetGluSerLysAspGlnGlyAla...LysAsnLeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31132 ATGGAGTCCAAAGAGACGCGTTAAACATCTCATCTGCGGAAATGT 31083

16 pHISGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLys 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31082 CAACACGAGAAATGATGAAAAGATGAAAAGGAGCAAGTTGCTAAATAAG 31033

33 rgGluProValAlaAlaProThrPheGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31032 GGGAGCCCTTGGCCCTACCT...TTGAATGTTAGTGAATCTGTGCCT 30986

50 ArgGlyGlyArgArgPheArgValArgGlnProIleSerHisTyra 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30985 AGAGGAATCGTAGGCGGTTCGCGTTAGCAGCCCATCTCGAGTATAG 30936

66 gTPAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGlu 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30935 ATGGGACATAATGCATAGGCTTGCAGAGCCACAGGCAAGGATCAGAGG 30886

83 LuAsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLys 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30885 AGAATATGAAAGGATTTGGGAGGAGGTGAGACAGCTGATGGAAGCTG 30836

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30835 AGGCAAAACAGTTGAGTCATAGTTTGGGGGAGTCAGCAGTATCCCC 30786

116 oHISHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30785 TCACCATGACCATCAGCATGAGTTTGGCTTATGCCCC 30749

seq_name: gb_hhg:HSU8081

```

```

seq_documentation_block:
LOCUS HSU8081 41029 bp DNA HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome X clone LL0XNC01-8081 map q22.1-22.3, ***

```







```
;
; TOPOLOGY: both
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 326...6277
US-08-836-325-9

alignment_scores:
  Quality: 72.50      Length: 58
  Ratio: 1.812       Gaps: 4
Percent Similarity: 68.966 Percent Identity: 36.207

alignment_block:
US-09-327-750D-34 x US-08-836-325-9/rev ..
Align seg 1/1 to reverse of: US-08-836-325-9 from: 1 to: 6452

25 LysAlaSerLysGlnSerGluGlu.....SerHisLeuG1 38
||||:||||| :|||: :|||:|
3220 AAGAGCCAGAAACAGAGTTTCAGGACCAAGGTTCCCAATCACCATGACCA 3171

38 uGluValGluAsnLysLysProGlyGlyAsnValArgArgLysValArgA 55
||||:||||| :|||: :|||:|
3170 TCATGTAACAATAAGACACATGTT.....CTGGCCCGGACCTC...C 3130

55 rgLeuValProAsnPhleuTriAlaIleProAsnArgHisValAspHis 71
||||:||||| :|||: :|||:|
3129 ATCAGTCCTCCACATGCTCTATCCACTCCACACAGCACTCGGAACAC 3080

72 SerGluGlyGly...GluGluVal 78
||||:| :|||:|
3079 GATGAGGAGGAGTGGAAGAAGTC 3056

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-892-770-1

seq_documentation_block:
; Sequence 1, Application US/08892770
; Patent No. 5962670
; GENERAL INFORMATION:
; APPLICANT: Walling, Linda L.
; APPLICANT: Pautot, Veronique
; APPLICANT: Gu, Yong-Qiang
; APPLICANT: Chao, Wun Shaw
; TITLE OF INVENTION: Improved promoters for Enhancing Plant
; TITLE OF INVENTION: Productivity
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/892,770
; FILING DATE: 15-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 023070-072100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 935 base pairs

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..935
; OTHER INFORMATION: /note= "tomato acidic leucine
; OTHER INFORMATION: aminopeptidase 1 (Lapal) promoter
; OTHER INFORMATION: region"
US-08-892-770-1

alignment_scores:
  Quality: 72.00      Length: 60
  Ratio: 2.250       Gaps: 2
Percent Similarity: 53.333 Percent Identity: 28.333

alignment_block:
US-09-327-750D-34 x US-08-892-770-1 ..
Align seg 1/1 to: US-08-892-770-1 from: 1 to: 935

62 TrpAlaIleProAsnArgHisValAspHisSerGluGlyGluGluVa 78
||||:||||| :|||: :|||:|
438 TGGTGTGTGTTGAATGAGATTATCGATCATATAAGAAAGAGGGGAGAAGAT 487

78 lGlyArgPheValGlyClnValMet.GluAlaLysArgHis..... 91
||||:||||| :|||: :|||:|
488 GAACAGTTTTTTTAAATAAATTTTGGCCCAAGAAAAAACATTCACGCGCT 537

92 .....SerLy 93
||||:|

538 GACTTTAAGTCATTACACTCACCTTCCCATGACAGAGAAAAATCGTCAAA 587

93 sGluGlnGlnMetArgProThrArg 102
||||:|
588 AAATGACACAATGAGACCCCTACATAAGG 615

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-726-214-5

seq_documentation_block:
; Sequence 5, Application US/08726214
; Patent No. 6107076
; GENERAL INFORMATION:
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gilman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,214
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,498
; FILING DATE: 04-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSD:450
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 4533 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; US-08-726-214-5

alignment_scores:
  Quality: 72.00      Length: 117
  Ratio: 1.075      Gaps: 5
Percent Similarity: 57.265      Percent Identity: 26.496

alignment_block:
US-09-327-750D-34 x US-08-726-214-5 ..
Align seg 1/1 to: US-08-726-214-5 from: 1 to: 4533

2 AlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAspLys 18
1918 GCATCCAGCCCGAGCTCCCTGCGCTTATTGAGACCAAGGAGCCCAATGG 1967
18 sLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluSerH 35
1968 GAGTGCCCATGCCAGCGGCTCCACATCAGAGGAGCTGAAGA..... 2010
35 isHisLeuGluGluValGluValAsnLysLysProGly...GlyAsnValArg 50
2011 .....CAGGAGGCCAGGCTGACAAACCCCTGTTCCCAACCCCGC 2052
51 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 67
2053 CGCAGGCTGCGC.....CTCCAGGACCTGGCAGACCG 2084
67 qHisValAspHisSerGluGlyGlyGluValGlyArgPheValGlyG 84
2085 TGTGTGGAGCGCTCTGAGGATGACGACGACTGACCACTGACCTCTTAACG 2134
84 lnValMetGluAlaLysArgHisSerLysGluGlnGlnMetArgProTyr 100
2135 AGGCCCTG...CTGGAGCGGAGTCCGCCCCAGGTGGTAAAGAGAGAAC 2181
101 Thr.....ArgPheArgThrProGluProAspAsnHisTy 112
2182 ACATTCTCTCAAGATGAGTGTTCATGGACCCAGAGATGGAACACGCTA 2231
112 r 112
2232 C 2232

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5180810-3

seq_documentation_block:
; Patent No. 5180810
; APPLICANT: Gomi, Hideyuki; Hozumi, Tatsunobu; Hattori, Shizuo;
; Tagawa, Chikaki; Kishimoto, Fumitaka; Bjorck, Lars
; TITLE OF INVENTION: PROTEIN H CAPABLE OF BINDING TO IGG
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/376,641
;   FILING DATE: 07-JUL-1989
; SEQ ID NO: 3:
;   LENGTH: 1131
; 5180810-3

alignment_scores:
  Quality: 71.00      Length: 100
  Ratio: 1.203      Gaps: 3
```

```
Percent Similarity: 59.000      Percent Identity: 20.000

alignment_block:
US-09-327-750D-34 x 5180810-3 ..
Align seg 1/1 to: 5180810-3 from: 1 to: 1131

3 SerLysValLysGlnValIleLeuAspLeuThrValGluLysAspLysLys 19
178 GCTAAACTTGTGAGGTGTGTGAACCAAGTTTGGAAACGAA..... 222
19 sasnLysLysGlyGlyLysAlaSerLysGlnSerGluGluSerHisH 36
223 .....AACTCAAGAGTGAGAAATGAGGAGCAATAAGAAAA 256
36 isLeuGluGluValGluAsnLysLysProGlyGlyAsnValArgArgLys 52
257 ATTTAGACAAACTTAGCAAAAGAAATCAAGGA.....AAG 291
53 ValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnArgHisVa 69
292 CTGCAAAATTTGAGCTTGACTATCTCAAAAATTTAGATCCGAGCAAA 341
69 lAspHisSerGluGlyGlyGluValGlyArgPheValGlyGlnValM 86
342 AGAGCACCAAAAGAACACACAGAA.....C 367
86 etLualalysArgHisSerLysGluGlnGlnMetArgProTyrThrArg 102
368 AAGAAGAGCGCACAAAAAATCAAGAACTTAGAAGCTTAATACCAACGA 417

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-045-583-6

seq_documentation_block:
; Sequence 6, Application US/09045583
; Patent No. 6287805
; GENERAL INFORMATION:
; APPLICANT: Graham, Gerard J. et al.
; TITLE OF INVENTION: No. 6287805el Molecules of the G Protein-Coupled
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
;   ADDRESS: LAHIVE & COCKFIELD, LLP
;   STREET: 28 State Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02109
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC Compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/045,583
;   FILING DATE: 20-MAR-98
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Mandragouras, Amy E.
;   REGISTRATION NUMBER: 36,207
;   REFERENCE/DOCKET NUMBER: MNI-044
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617)227-7400
;   TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1137 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: cDNA
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1134
US-09-045-583-6

alignment_scores:
  Quality: 71.00      Length: 114
  Ratio: 1.479       Gaps: 4
Percent Similarity: 42.105 Percent Identity: 22.807

alignment_block:
US-09-327-750D-34 x US-09-045-583-6/rev ..
Align seg 1/1 to reverse of: US-09-045-583-6 from: 1 to: 1137

      4 LysValLysGlnValIleLeuAspLeuThrValGluLysAsp..... 17
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      876 CGTATAGTCAGACGGTGGCTGATCTCACAGTTCCCAAGACATGCAGGT 827
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      18 ....LysLysAsnLysGlyGlyLysAla.....SerLysGlnSerg 31
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      826 CCAGCACACCATCTAAGGTGAGGTGAGGTATATATGGGAACACAGCATG 777
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      31 luGluGluSerHis..... 35
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      776 AAGAAACTATCACAGGGCGGCGGCATCCTCAGAGCCGCGCCTGGCC 727
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      35 ..... 35
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      726 TGGCGGCGCTCAGCCTGACCAGAACGCAACCGATGCGGGAGTAAAGAAGA 677
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      36 .HisLeuGluGluValGluAsnLysLys..... 44
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      676 TCATGGCCAAAGAGTGGGAGGAGAAACCCAGAAAGTTTCAGCTGGAAGCGC 627
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      45 .....ProGlyGlyAsnValArgArgLysValArgLysVal 57
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      626 AGGTACAGCTTCCAAATGGTCGATGTCGCCCAAAATCCGCATAGCAGTG 577
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      58 ProAsnPheLeuTrpAlaIleProAsnArgHisValAspHis 71
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      576 CCACACACCATCTAAGGTGAGGTGAGGTGATCTGCACAAAGACCAT 535
      : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-045-583-4

seq_documentation_block:
; Sequence 4, Application US/09045583
; Patent No. 6287805
; GENERAL INFORMATION:
; APPLICANT: Graham, Gerard J. et al.
; TITLE OF INVENTION: No. 6287805el Molecules of the G Protein-Coupled
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045.583
; FILING DATE: 20-MAR-98
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MNI-044
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-4214
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 193..1326
US-09-045-583-4

alignment_scores:
  Quality: 71.00      Length: 114
  Ratio: 1.479       Gaps: 4
Percent Similarity: 42.105 Percent Identity: 22.807

alignment_block:
US-09-327-750D-34 x US-09-045-583-4/rev ..
Align seg 1/1 to reverse of: US-09-045-583-4 from: 1 to: 1664

      4 LysValLysGlnValIleLeuAspLeuThrValGluLysAsp..... 17
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      1068 CGTATAGTCAGACGGTGGCTGATCTCACAGTTCCCAAGACATGCAGGT 1019
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      18 ....LysLysAsnLysGlyGlyLysAla.....SerLysGlnSerg 31
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      1018 CCAGCAACGAGTGCAGAAACAGGTGAGGTATATATGGGAACACAGCATG 969
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      31 luGluGluSerHis..... 35
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      968 AAGAAACTATCACAGGGCGGCGGCATCCTCAGAGCCGCGCCTGGCC 919
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      35 ..... 35
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      918 TGGCGGCGCTCAGCCTGACCAGAACGCAACCGATGCGGGAGTAAAGAAGA 869
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      36 .HisLeuGluGluValGluAsnLysLys..... 44
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      868 TCATGGCCAAAGAGTGGGAGGAGAAACCCAGAAAGTTTCAGCTGGAAGCGC 819
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      45 .....ProGlyGlyAsnValArgArgLysValArgLysVal 57
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      818 AGGTACAGCTTCCAAATGGTCGATGTCGCCCAAAATCCGCATAGCAGTG 769
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      58 ProAsnPheLeuTrpAlaIleProAsnArgHisValAspHis 71
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      768 CCACACACCATCTAAGGTGAGGTGAGGTGATCTGCACAAAGACCAT 727
      : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5180810-2

seq_documentation_block:
; Patent No. 5180810
; APPLICANT: Gomi, Hideyuki;Hozumi, Tatsunobu;Hattori, Shizuo;
; Tagawa, Chiaki;Kishimoto, Fumitaka;Bjorek, Lars
; TITLE OF INVENTION: PROTEIN H CAPABLE OF BINDING TO IGG
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/376.641
; FILING DATE: 07-JUL-1989
; SEQ ID NO:2:
; LENGTH: 1784
5180810-2

```

alignment\_scores:  
Quality: 71.00 Length: 100  
Ratio: 1.203 Gaps: 3  
Percent Similarity: 59.000 Percent Identity: 20.000

alignment\_block:

US-09-327-750D-34 x 5180810-2 ..

Align seg 1/1 to: 5180810-2 from: 1 to: 1784

```
3 SerLysValLysGlnValLleLeuAspLeuThrValGluLysAspLysLy 19
505 GCTAAATCTTGTAGGTTGTGAAACACACAAAGTTTGGAAACGAA..... 549
19 sAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGluSerHisH 36
550 .....AAACTCAAGAGTGAGATGAGGAGCAATGAAGAA 583
36 IsLeuGluGluValGluAsnLysLysProGlyGlyAsnValArgArgLys 52
584 ATTTAGACAAACTTAGCAAGAAATCAAGGA.....AAG 618
53 ValArgArgLeuValProAsnPhLeuTrpAlaLleProAsnArgHisVa 69
619 CTCGAAAAATGTCAGCTTGACTATCTCAAAAAAATTAGATCAGCAGCACAA 668
69 lAspHisSerGluGlyGlyGluGluValGlyArgPheValGlyGlnValM 86
569 AGAGCACCAACCAAGAACACAA.....C 694
86 etGluAlaLysArgHisSerLysGluGlnGlnMetArgProTyrThrArg 102
695 AAGAAGAGCGCACAAAAAATCAAGAACAAATTAGACGCTAAATACCAACGA 744
```

seq\_name: /cgn2.6/ptodata/2/lna/5B\_COMB.seq:US-08-841-483-3

seq\_documentation\_block:

; Sequence 3, Application US/08841483B

; Patent No. 5976875

; GENERAL INFORMATION:

; APPLICANT: Prescott, Steven M.

; APPLICANT: Bunting, Michaeline

; APPLICANT: Tang, Wen

; APPLICANT: Topham, Matthew

; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and

; FILE REFERENCE: 2037.2.1a

; CURRENT APPLICATION NUMBER: US/08/841,483B

; EARLIER FILING DATE: 1997-04-22

; EARLIER FILING DATE: 1996-04-22

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; TYPE: DNA

; LENGTH: 3490

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (89)..(2875)

US-08-841-483-3

alignment\_scores:  
Quality: 71.00 Length: 82  
Ratio: 1.651 Gaps: 4  
Percent Similarity: 52.439 Percent Identity: 30.488

alignment\_block:

US-09-327-750D-34 x US-08-841-483-3 ..

Align seg 1/1 to: US-08-841-483-3 from: 1 to: 3490

```
13 ThrValGluLysAspLysLysAsnLysLysGlyGly.....LysAl 26
|||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
848 ACTCTGAAAGCAAGCAAGAAAGAAAGAGAGGCGCATCTTCAAGAGGAAGTC 857
26 aSerLysGlnSerGluGluGluSerHis..... 35
|||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
898 CAGCAAGAAAGGCGCTGAGGAGGGCGCTGGAGACCCCTTCATCATCAGGC 947
36 .....HisLeuGluGluValGluAsnLysLys 44
948 CCACCCCTCCCGCTCATGAAGCCCTGCTGGTGTGTGTGAACCCCAAG 997
45 ProGlyGlyAsnValArgArgLysValArgLeuValProAsnPhLe 61
|||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
998 AGTGGGGCAACCAAGGTCGAAGATC.....ATCCAGTCTTTCCT 1038
61 uTrpAlaLleProAsnArgHisVal...AspHisSerGluGlyGly 75
|||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
1039 CTGGTATCTCAATCCCGCACAAAGTCTTCGACCTGAGCCAGGAGGG 1084
```

seq\_name: /cgn2.6/ptodata/2/lna/6B\_COMB.seq:US-09-382-911-3

seq\_documentation\_block:

; Sequence 3, Application US/09382911

; Patent No. 6221658

; GENERAL INFORMATION:

; APPLICANT: Prescott, Steven M.

; APPLICANT: Bunting, Michaeline

; APPLICANT: Tang, Wen

; APPLICANT: Topham, Matthew

; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and

; FILE REFERENCE: 2037.2.1a

; CURRENT APPLICATION NUMBER: US/09/382,911

; PRIOR FILING DATE: 1999-08-25

; PRIOR FILING DATE: 1997-04-22

; PRIOR FILING DATE: 1996-04-22

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; TYPE: DNA

; LENGTH: 3490

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (89)..(2875)

US-09-382-911-3

alignment\_scores:

Quality: 71.00 Length: 82

Ratio: 1.651 Gaps: 4

Percent Similarity: 52.439 Percent Identity: 30.488

alignment\_block:

US-09-327-750D-34 x US-09-382-911-3 ..

Align seg 1/1 to: US-09-382-911-3 from: 1 to: 3490

```
13 ThrValGluLysAspLysLysAsnLysLysGlyGly.....LysAl 26
|||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
848 ACTCTGAAAGCAAGCAAGAAAGAAAGAGGCGCATCTTCAAGAGGAAGTC 897
26 aSerLysGlnSerGluGluGluSerHis..... 35
|||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
898 CAGCAAGAAAGGCGCTGAGGAGGGCGCTGGAGACCCCTTCATCATCAGGC 947
36 .....HisLeuGluGluValGluAsnLysLys 44
948 CCACCCCTCCCGCTCATGAAGCCCTGCTGGTGTGTGTGAACCCCAAG 997
```

```

; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; APPLICANT: Topham, Matthew
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; TITLE OF INVENTION: Methods of Use Thereof
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/09/382,911
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 08/841,483
; PRIOR FILING DATE: 1997-04-22
; PRIOR APPLICATION NUMBER: 60/016,210
; PRIOR FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 5
; LENGTH: 4094
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (126)..(3479)
US-09-382-911-5

alignment_scores:
      Quality: 71.00      Length: 82
      Ratio: 1.651      Gaps: 4
Percent Similarity: 52.439      Percent Identity: 30.488

alignment_block:
US-09-327-7500-34 x US-09-382-911-5 ..

Align seg 1/1 to: US-09-382-911-5 from: 1 to: 4094

13 ThrValGluLysAspLysLysAsnLysLysGlyGly.....LysAl 26
|||||: :|||:|||||:|||||:|||||:|||||:
1452 ACTCTGAAGCAAGCAAGAGAGAGAGAGAGGCGATCTTCAAGGAAGTC 1501

26 aserLysclserclugluGluSerHis..... 35
|||||:|||||:|||||:|||||:|||||:
1502 CAGCAAGAAAGGCGCTGAGGAGGCGCGTGAGACCTTCATCATCAGGC 1551

36 .....HisLeuGluGluValGluAsnLysLys 44
|||||:|||||:|||||:|||||:
1552 CCACCCCTCCCGCTCATGAGCCCTGCTGGTGTGTTGTGAACCCCAAG 1601

45 ProGlyGlyAsnValArgArgLysValArgArgLeuValProAsnPheLe 61
|||||:|||||:|||||:|||||:|||||:
1602 AGTGGGGCAACCAAGGCTGCAAGATC.....ATCCAGTCTTTCTCT 1642

61 uTPAlaIleProAsnArgHisVal...AspHisSerGluGlyGly 75
|||||: :|||:|||||:|||||:|||||:
1643 CTGTATCTATCAATCCCGACCAAGTCTTCGACCTGAGCCAGGGAGG 1688

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-807-043B-5

seq_documentation_block:
; Sequence 5, Application US/07807043B
; Patent No. 5342774
; GENERAL INFORMATION:
; APPLICANT: Boon, Thierry, Van den Eynde, Beno t
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

```

```
;
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,043B
; FILING DATE: 19911212
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; FILING DATE: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5342774man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 253.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: singular
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-07-807-043B-5

alignment_scores:
    Quality: 71.00      Length: 93
    Ratio: 1.365        Gaps: 4
    Percent Similarity: 55.914      Percent Identity: 27.957

alignment_block:
US-09-327-750D-34 x US-07-807-043B-5/rev ..

Align seg 1/1 to reverse of: US-07-807-043B-5 from: 1 to: 4698

16 LysAspLysLysAsnLysGlyGlyLysAlaSer.LysGlnSerGluG 32
2134 AGGCAGGAAAGCAGCAGGAGGGGAGCCAGCGAGGGGAGGAC 2085
32 luGluSerHisLysLeuGluGluValGluAsnLys.....Lys 44
2084 AGGAAAGGCACACAGAGGGGAGCGGAGGGGGGATCTTGGAAA 2035
45 ProGlyGlyAsnValArgArgLysValArgArgLeuValProAsnPheLe 61
2034 CAGGTGCAAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1990
61 utrAlaIleProAsnArgHisValAspHisSerGluGlyGluGluV 78
1989 .....CCAAACAACAAAAAGGTGAGCTGGAAGGTGAGGAGGAGGAGC 1950
78 alGly.....ArgPheValGlyGlnValMetGluAlaLys 89
1949 CAGGAGGAAAGGCACAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGT 1900
90 ArgHisSerLysGluGlnGlnMetArg 98
1899 AGCGTGTCTAAAGTCGACGCAAGAGG 1873

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-299-849B-5

seq_documentation_block:
; Sequence 5, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:
; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
; APPLICANT: Leth , Bernard; Szikora, Jean-Pierre; De Smet, Charles;
; APPLICANT: Chomez, Patrick
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
; DETERMINING EXPRESSION OF A Tumor Antigen Precursor
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5612201man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-299-849B-5

alignment_scores:
    Quality: 71.00      Length: 93
    Ratio: 1.365        Gaps: 4
    Percent Similarity: 55.914      Percent Identity: 27.957

alignment_block:
US-09-327-750D-34 x US-08-299-849B-5/rev ..

Align seg 1/1 to reverse of: US-08-299-849B-5 from: 1 to: 4698

16 LysAspLysLysAsnLysGlyGlyLysAlaSer.LysGlnSerGluG 32
2134 AGGCAGGAAAGCAGCAGGAGGGGAGCGGAGGGGAGGAC 2085
32 luGluSerHisLysLeuGluGluValGluAsnLys.....Lys 44
2084 AGGAAAGGCACACAGAGGGGAGCGGAGGGGGGATCTTGGAAA 2035
```

**THIS PAGE BLANK (USPTO)**

```

45 ProGlyGlyValArgArgLysValArgArgLeuValProAsnPheLe 61
   |||:||||: |||:||||:||||:||||: |||: |||: |||:
2034 CAAGGTGCRAAAAAAAAAAAGCAACCAACCAACCAACCAACAA 1990
   |||:||||: |||:||||: |||: |||: |||: |||: |||:
61 uTrpAlaIleProAsnArgHisValAspHisSerGlyGlyGluGluV 78
   |||:||||: |||:||||: |||: |||: |||: |||: |||:
1989 .....CCAAACAACAAAAAGGTGAGCTGGAAGGTGGAGGGGAGC 1950
   |||:||||: |||:||||: |||: |||: |||: |||: |||:
78 aGly.....ArgPheValGlyGlnValMetGluAlaLys 89
   |||:||||: |||:||||: |||: |||: |||: |||: |||:
1949 CAGGAGGAAGGCACAAAAAAGGGGAACGGGGCAGAAAGCAGGGAGAGT 1900
   |||:||||: |||:||||: |||: |||: |||: |||: |||:
90 ArgHisSerLysGluGlnGlnMetArg 98
   |||:||||: |||:||||: |||: |||: |||: |||: |||:
1899 AAGGCTGCTAAAGTCGAGCAAGAGG 1873
```



OM of: US-09-327-750D-34 to: N\_Geneseq\_1101.\* out\_format : pfs  
Date: Mar 11, 2002 3:42 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09327750/runat\_11032002\_101154\_20362/app\_query.fasta\_1.1472  
-DB=N\_Geneseq\_1101 -OPMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINWATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blousum62  
-TRANS=human4.0.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pfs  
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09327750@cgn1\_1.330 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-34

Query length: 118

Database: N\_Geneseq\_1101.\*

Database sequences: 930621

Database length: 428662619

Search time (sec): 355.560000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH03517		372.00	800.02	865	1.7e-36
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH13750		372.00	796.41	1229	2.8e-36
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH03548		372.00	795.34	1364	3.2e-36
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF23529		171.50	358.04	891	7.2e-12
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAF21748		174.00	354.44	917	1.1e-11
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF23528		140.00	299.94	700	1.2e-08
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH01004		143.00	301.51	485	1.0e-08
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH01005		136.00	285.14	532	8.3e-08
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH75810		111.00	226.31	767	0.0002
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH58581		111.00	225.10	862	0.0002
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF59611		111.00	224.68	898	0.0002
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH45143		109.50	222.67	792	0.0003
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH03880		103.00	210.20	662	0.0012
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH03879		103.00	209.65	698	0.0013
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH160367		99.00	198.72	858	0.0054
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH010889		98.50	199.83	692	0.0047
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH01003		97.50	203.23	401	0.0030
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH23519		92.50	143.81	6.20	44453
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH20201		89.50	189.45	0.0178	276
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH15401		89.50	189.45	0.0178	276
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH15905		89.50	189.45	0.0178	276
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH13308		87.50	180.85	0.0536	415
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH134661		87.50	180.85	0.0536	415
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH03209		87.50	180.85	0.0536	415
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH257839		83.50	158.39	0.9561	1566
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH10225		83.50	154.98	250.27	106746
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH08625		81.00	150.02	2.80	2069
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH15484		80.50	159.18	0.8631	762
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH28355		80.00	114.62	262.02	52216
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH46153		78.00	137.19	14.49	3786
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH77225		78.00	136.98	14.89	3866
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH08642		78.00	136.82	15.19	3925
/SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:AAH086237		78.00	135.42	18.19	4500
/SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:AAH086236		78.00	135.31	18.45	4548
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH70222		77.50	131.37	30.58	5994
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH39709		77.00	150.83	2.52	811
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH22774		77.00	147.68	3.77	1102
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH121493		76.00	156.05	1.29	394
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH46784		76.00	156.05	1.29	394
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH07190		76.00	156.05	1.29	394

/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH54129 + 76.00 133.47 23.35  
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH93013 - 75.50 149.04 3.17  
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH73887 - 75.50 142.27 7.55  
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH72409 + 75.00 142.42 7.41  
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH00010 - 75.00 133.35 23.71  
289:  
seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH03517  
seq\_documentation\_block:  
ID AAH03517 standard; cDNA; 865 BP.  
XX  
AC AAH03517;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA clone (5'-primer) SEQ ID NO:352.  
DE  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
XX  
XX 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs.  
XX  
XX Claim 1; SEQ ID 352; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
full-length cDNAs defined in the specification. Where a primer set  
comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
to the complementary strand of a polynucleotide which comprises one of  
the 5602 nucleotide sequences defined in the specification, where the  
oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
of an oligonucleotide comprising a sequence complementary to the  
complementary strand of a polynucleotide which comprises a 5'-end  
sequence and an oligonucleotide comprising a sequence complementary to a  
polynucleotide which comprises a 3'-end sequence, where the  
oligonucleotide comprises at least 15 nucleotides and the combination of  
the 5'-end sequence/3'-end sequence is selected from those defined in  
the specification. The primer sets can be used in antisense therapy and  
in gene therapy. The primers are useful for synthesizing polynucleotides,  
particularly full-length cDNAs. The primers are also useful for the  
detection and/or diagnosis of the abnormality of the proteins encoded by  
the full-length cDNAs. The primers allow obtaining of the full-length  
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
represent oligonucleotides, all of which are used in the exemplification  
of the present invention.

Sequence 865 BP; 239 A; 177 C; 224 G; 221 T; 4 other;

```
alignment_scores:
  Quality: 372.00      Length: 120
  Ratio: 3.758        Gaps: 2
  Percent Similarity: 82.500      Percent Identity: 60.000

alignment_block:
  US-09-327-750D-34 x AAH03517 ..

  Align seg 1/1 to: AAH03517 from: 1 to: 865

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17
  ||| ||||| : : : : : ||||| : : : : : ||||| : : : : :
191 ATGGAGTCCAAAGAGAACTAGCGGCAACAATCTCAACGGGAAATGC 240

17 pLysLysAsnLysGlyGly...LysAlaSerLysGlnSerGluGluG 33
  : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 CCACAGAAACCAAGAGGGGAGCGAGCCGCCACGCAAGTAAGAAG 290

33 LuSerHisHisValAspHisSerGlyGlyGlnValLysProGlyGlyAsnVal 49
  ||||| : : : : : ||||| : : : : : ||||| : : : : :
291 AATCCCGCATTTGGAGGGGTCAAGCCAGAGCCCTGGAGGAATATC 340

50 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 65
  ||||| : : : : : ||||| : : : : : ||||| : : : : :
341 AGCGGGGGCGGAGTTCAGCGACTTGTCCCTAAATTTTCGATGGGCATACC 390

65 oAsnArgHisValAspHisSerGlyGlyGlnValLysGlyArgPheV 82
  ||||| : : : : : ||||| : : : : : ||||| : : : : :
391 TAATAGCATATTGAGCACATGAGCGAGAGATGATGTAGAAAGTTTG 440

82 alGlyGlnValMetGluAlaLysArgHisSerLysGlnGlnMetArg 98
  ||||| : : : : : ||||| : : : : : ||||| : : : : :
441 TAGGCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATGAGG 490

99 ProTyrThrArgPheArgThrProGluProAsnHisTyrAspPheCy 115
  ||| ||||| : : : : : ||||| : : : : : ||||| : : : : :
491 CACTATATGCGCTTCCAACTCTGAACTGACCAACCATTTATGACTTTTG 540

115 sLeuIlePro 118
  ||||| |||||
541 CCTCATACCT 550
```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH13750

seq\_documentation\_block:

ID AAH13750 standard; cDNA; 1229 BP.

XX AC AAH13750;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA sequence SEQ ID NO:10656.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

```
XX WPI: 2001-318749/34.
XX DR
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs .
XX
PS Claim 8; SEQ ID 10656; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC in the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1229 BP; 351 A; 243 C; 296 G; 339 T; 0 other;
```

alignment\_scores:

Quality: 372.00 Length: 120

Ratio: 3.758 Gaps: 2

Percent Similarity: 82.500 Percent Identity: 60.000

alignment\_block:

US-09-327-750D-34 x AAH13750 ..

Align seg 1/1 to: AAH13750 from: 1 to: 1229

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17

||| ||||| : : : : : ||||| : : : : : ||||| : : : : :

191 ATGGAGTCCAAAGAGAACTAGCGGCAACAATCTCAACGGGAAATGC 240

17 pLysLysAsnLysGlyGly...LysAlaSerLysGlnSerGluGluG 33

: :

241 CCACAGAAACCAAGAGGGGAGCGAGCCGCCACGCAAGTAAGAAG 290

33 LuSerHisHisValAspHisSerGlyGlyGlnValLysProGlyGlyAsnVal 49

||||| : : : : : ||||| : : : : : ||||| : : : : :

291 AATCCCGCATTTGGAGGGGTGAAGCCAGAGCCCTGGAGGAATATC 340

50 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 65

||||| : : : : : ||||| : : : : : ||||| : : : : :

341 AGCGGGGGCGGAGTTAGCGACTTGTCCCTAAATTTTCGATGGGCCATACC 390

65 oAsnArgHisValAspHisSerGlyGlyGlnValLysGlyArgPheV 82

||||| : : : : : ||||| : : : : : ||||| : : : : :

391 TAATAGCATATTGAGCACATGAGCGAGAGATGATGTAGAAAGTTTG 440

82 alGlyGlnValMetGluAlaLysArgHisSerLysGlnGlnMetArg 98

||||| : : : : : ||||| : : : : : ||||| : : : : :

441 TAGGCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATGAGG 490

99 ProTyrThrArgPheArgThrProGluProAsnHisTyrAspPheCy 115

||| ||||| : : : : : ||||| : : : : : ||||| : : : : :

491 CACTATATGCGCTTCCAACTCTGAACTGACCAACCATTTATGACTTTTG 540





486 GATATGAATATTTCATGGAGGAGATCAGAGAAATCAGAGAAACTTAG 535  
93 sGluGlnMetArgProTyrThrArgPheArgThrProGluProAspA 110  
536 GGAGCTGCAGTTGAGGAATGCTGCGTATCCTTATGGGGGAGCTCTCTA 585  
110 snHisTyrAsp.....PheCysLeuIlePro 118  
586 ATCACCATGACCATCATGATGAATTTTGCCTTATGCCT 623

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF23528

seq\_documentation\_block:

ID AAF23528 standard; DNA; 700 BP.

AC AAF23528;

DT 22-MAR-2001 (first entry)

DE Mouse NADE DNA.

KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;  
KW neurogenetic disease; NF-kappaB; ds.

OS Mus sp.

PN W0200075278-A2.

PD 14-DEC-2000.

PF 07-JUN-2000; 2000WO-US15621.

PR 07-JUN-1999; 99US-0327750.

PA (UYCO ) UNIV COLUMBIA NEW YORK.

PI Sato T;

DR WPI; 2001-061707/07.

PT New p75-neurotrophin receptor-associated cell death executor (NADE) and  
PT the gene encoding NADE, useful for modulating the activity of p75NTR  
PT and for detecting neurodegenerative diseases .

PS Claim 12; Fig 1; 134pp; English.

CC The present invention relates to a purified polypeptide capable of  
CC binding neurotrophin receptor (p75-NTR). The invention is useful for  
CC binding and modulating the activity of p75NTR. The peptide mediates  
CC NGF-induced apoptosis, which plays an important role in neurogenetic  
CC diseases. The peptide of the invention and p75NTR are useful for  
CC inhibiting NF-kappaB activation in a cell or a subject, for inducing  
CC caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase  
CC and fragment nuclear DNA in a cell by co-expression of (I) and p75-NTR.

SQ Sequence 700 BP; 177 A; 188 C; 203 G; 132 T; 0 other;

alignment\_scores:

Quality: 144.00 Length: 129  
Ratio: 2.000 Gaps: 3  
Percent Similarity: 55.814 Percent Identity: 28.682

alignment\_block:

US-09-327-750D-34 x AAF23528 ..

Align seg 1/1 to: AAF23528 from: 1 to: 700

18 LysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlucose 34  
162 AAACAAATCTCATGTCGCAATGTCCACGAGAAACGAGAGCTGGA 211  
34 rHisHisLeuGluGluValGluAsnLysLysProGlyGly..... 47

212 GCAGCCCTGCAGATGGACAGGAAACCCCTGTGGGAGGAGGTGAGG 261  
47 .....  
262 GCACACAGCCTGCTCAAAACAACAACAACAACAACAACACCAACACCA 311  
48 .....AsnValArgArgLysValArgArgLeuValProAsnPheLeuTr 62  
312 ACCACACCGAGAGAGCCAGGCTCGCCGACTTGGCCCTAACTTCCGATG 361  
62 pAlaIleProAsnArgHisValAspHisSerGluGly.....GlyGluG 77  
362 GGCCATTCCACACAGCAGATGAATCAGCGGTTGGGTGGAGATGGAGATG 411  
77 luValGlyArgPheValGlyGlnValMetGluAlaLysArgHisSerLys 93  
412 ATATGGAATGTTCATGGAGGAGATGAGAGATCCGGAGAAAGCTTAGG 461  
94 GluGlnGlnMetArgProTyrThrArgPheArgThrProGluProAspAs 110  
462 GAGCTACAGCTGAGAAATGCTCTACGCATCCTTATGGGGGAGCTGTCTAA 511  
110 nHisTyrAsp.....PheCysLeuIlePro 118  
512 CCACACGATCACCATGATGAATTCCTGCTTATGCCT 548

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA01004

seq\_documentation\_block:

ID AA01004 standard; cDNA; 485 BP.

AC AA01004;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 1002.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

DR P-PSDB; AAG00998.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 1002; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from











```
XX
SQ   Sequence 662 BP; 177 A; 144 C; 192 G; 137 T; 12 other;

alignment_scores:
    Quality: 103.00      Length: 132
    Ratio: 1.411        Gaps: 6
    Percent Similarity: 55.303      Percent Identity: 30.303

alignment_block:
US-09-327-750D-34 x AAC03880 ..
Align seg 1/1 to: AAC03880 from: 1 to: 662
1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrVal..... 14
207 ATGGAGTCCAAAGAGAACGAGCGTTAAACAATCTCATCGTGGAATGT 256
15 .....GluLysAspLysLysAsnLysLysGlyGlyLysAlaSerLysG 29
257 CAACGAGAAATGATGAAAGATGAAAGAGCAAGTGWGCTAATAAA.. 305
29 lnSerGluGluGluSerHisHisLeuGluGluValGluAsnLysLysPro 45
306 ..GGGAGCCCTTGGCCCTACCTTTGATGCTGTGTGCTGCTGCT 353
46 GlyGlyAsnValArgArg...LysValArgArgLeuValProAsnPheLe 61
354 AGAGGAATTCGTAGGCGGTTCCGCGTTAGCGACCCATCTCGCAGTATAG 403
61 uTrpAlaIleProAsnArgHisValAsp.....HisSerG 73
404 ATGGATATGATGCATAGGCTTGGAGACCAAGCAAGGATGACAGAGAG 453
73 luGlyGlyGluGluValGlyArgPheValGlyGlnValMetGluAlaLys 89
454 AGAATATGGAAGAGATTGGGAGGAGGTGACAGAGCTGATGGAAAAAGCTG 503
90 ArgHisSerLysGluGlnGlnMetArgProTyrThrArgPheArgThrPr 106
504 AGG.....GAAAGCAGTTGAGTCATAGTCTCGGGGCGAGTCAGCAC 544
106 oGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 118
545 TGACCCCTCCACCATGACCATCATGATGATGATGATGATGATGATGATG 590
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAC03879
seq_documentation_block:
ID AAC03879 standard; cDNA; 698 BP.
XX
AC AAC03879;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3877.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
```

```
XX
DR WPI: 2000-500381/45.
P-PSDB; AAG03873.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 3877; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 698 BP; 187 A; 151 C; 210 G; 139 T; 11 other;

alignment_scores:
    Quality: 103.00      Length: 132
    Ratio: 1.411        Gaps: 6
    Percent Similarity: 55.303      Percent Identity: 30.303

alignment_block:
US-09-327-750D-34 x AAC03879 ..
Align seg 1/1 to: AAC03879 from: 1 to: 698
1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrVal..... 14
243 ATGGAGTCCAAAGAGAACGAGCGTTAAACAATCTCATCGTGGAATGT 292
15 .....GluLysAspLysLysAsnLysLysGlyGlyLysAlaSerLysG 29
293 CAACGAGAAATGATGAAAGATGAAAGAGCAAGTGWGCTAATAAA.. 341
29 lnSerGluGluGluSerHisHisLeuGluGluValGluAsnLysLysPro 45
342 ..GGGAGCCCTTGGCCCTACCTTTGATGCTGTGTGCTGCTGCTGCT 389
46 GlyGlyAsnValArgArg...LysValArgArgLeuValProAsnPheLe 61
390 AGAGGAATTCGTAGGCGGTTCCGCGTTAGCGACCCATCTCGCAGTATAG 439
61 uTrpAlaIleProAsnArgHisValAsp.....HisSerG 73
440 ATGGATATGATGCATAGGCTTGGAGACCAAGCAAGGATGACAGAGAG 489
73 luGlyGlyGluGluValGlyArgPheValGlyGlnValMetGluAlaLys 89
490 ACAATATGGAAGAGATTGGGAGGAGGTGACAGACAGCTGATGCAAGAGCTG 539
90 ArgHisSerLysGluGlnGlnMetArgProTyrThrArgPheArgThrPr 106
540 AGG.....GAAAGCAGTTGATGCTCATAGTCTCGGGGCGAGTCAGCAC 580
106 oGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 118
581 TGACCCCTCCACCATGACCATCATGATGATGATGATGATGATGATGATG 626
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI60367
seq_documentation_block:
```

**THIS PAGE BLANK (USPTO)**







details see [http://bacpac.med.buffalo.edu/ VECTOR:pcypac2](http://bacpac.med.buffalo.edu/VECTOR:pcypac2)  
 IMPORTANT: This sequence is not the entire insert of clone 635G19.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we arrange for a small overlap between  
 neighbouring submissions.  
 The true left end of clone U65A4 (Z81014) is at 69549 in this  
 sequence. The true right end of clone U101D3 (Z85997) is at 100 in  
 this sequence.

## FEATURES

Location/Qualifiers  
 source

```

1..69648
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q22.1-22.3"
/clone="RP4-635G19"
/clone.lib="RPCI-4"
439..483
repeat_region
/note="TAR1 repeat: matches 707..749 of consensus"
501..661
repeat_region
/note="AluYb repeat: matches 137..311 of consensus"
662..897
repeat_region
/note="118 copies 2 mer aa 70% conserved"
915..1331
repeat_region
/note="MSTR repeat: matches 1..419 of consensus"
1334..1517
repeat_region
/note="92 copies 2 mer ga 77% conserved"
1927..2020
repeat_region
/note="MIR repeat: matches 109..217 of consensus"
2066..2262
repeat_region
/note="L1M2 repeat: matches 5825..6022 of consensus"
2263..2574
repeat_region
/note="AluSp repeat: matches 1..313 of consensus"
2575..2819
repeat_region
/note="L1M2 repeat: matches 5583..5825 of consensus"
2821..3039
repeat_region
/note="Alu repeat: matches 1..311 of consensus"
3104..3384
repeat_region
/note="L1M2 repeat: matches 5304..5591 of consensus"
3391..4007
repeat_region
/note="L1M1 repeat: matches 5558..6162 of consensus"
4012..4137
repeat_region
/note="Trigger2a repeat: matches 11..123 of consensus"
4141..4252
repeat_region
/note="L1 repeat: matches 2920..3031 of consensus"
4265..4350
repeat_region
/note="HY4 repeat: matches 1..87 of consensus"
4350..4678
repeat_region
/note="MER74A repeat: matches 28..369 of consensus"
4914..5095
repeat_region
/note="MER5A repeat: matches 8..189 of consensus"
5224..5555
repeat_region
/note="L2 repeat: matches 2423..2746 of consensus"
6040..6086
repeat_region
/note="L2 repeat: matches 1676..1720 of consensus"
7413..7478
repeat_region
/note="Alu repeat: matches 239..304 of consensus"
7628..7928
repeat_region
/note="AluYb8 repeat: matches 1..316 of consensus"
7952..7987
repeat_region
/note="18 copies 2 mer aa 81% conserved"
7990..8305
repeat_region
/note="AluJo repeat: matches 1..309 of consensus"
8436..8538
repeat_region
/note="MIR repeat: matches 99..212 of consensus"
10836..11442
repeat_region
/note="L1M2 repeat: matches 1357..1972 of consensus"
11443..11571
repeat_region
/note="FLAM.A repeat: matches 2..129 of consensus"
11572..11946
repeat_region
/note="L1M2 repeat: matches 977..1357 of consensus"
11947..12245
repeat_region
/note="AluY repeat: matches 1..301 of consensus"
12246..13098
repeat_region

```

```

repeat_region
/note="L1M2 repeat: matches 275..977 of consensus"
13094..13223
repeat_region
/note="L1M2 repeat: matches 17..113 of consensus"
13523..13820
repeat_region
/note="MLTID repeat: matches 177..505 of consensus"
13823..13932
repeat_region
/note="55 copies 2 mer ct 77% conserved"
13936..13977
repeat_region
/note="MLTID repeat: matches 201..241 of consensus"
13978..14240
repeat_region
/note="AluSg1 repeat: matches 1..270 of consensus"
14245..14534
repeat_region
/note="AluSc repeat: matches 1..290 of consensus"
14535..14721
repeat_region
/note="MLTID repeat: matches 2..201 of consensus"
14724..15604
repeat_region
/note="L1MC/D repeat: matches 4646..5565 of consensus"
15605..15909
repeat_region
/note="AluSc repeat: matches 24..312 of consensus"
16069..16363
repeat_region
/note="AluX repeat: matches 1..291 of consensus"
16482..16641
repeat_region
/note="AluYb repeat: matches 136..295 of consensus"
16671..16953
repeat_region
/note="AluSc repeat: matches 1..284 of consensus"
17164..17280
repeat_region
/note="L1P repeat: matches 4499..4619 of consensus"
17326..17377
repeat_region
/note="26 copies 2 mer at 79% conserved"
17704..17863
repeat_region
/note="AluSc repeat: matches 118..295 of consensus"
17864..18175
repeat_region
/note="AluYb8 repeat: matches 1..318 of consensus"
18176..18298
repeat_region
/note="AluSc repeat: matches 1..118 of consensus"
18848..18873
repeat_region
/note="13 copies 2 mer tg 100% conserved"
18876..19171
repeat_region
/note="AluSg repeat: matches 1..295 of consensus"
19183..19232
repeat_region
/note="25 copies 2 mer ac 100% conserved"
19246..19533
repeat_region
/note="AluY repeat: matches 5..292 of consensus"
19539..19837
repeat_region
/note="MER74B repeat: matches 61..368 of consensus"
20016..20194
repeat_region
/note="MER74A repeat: matches 295..487 of consensus"
20394..20692
repeat_region
/note="AluSg repeat: matches 1..302 of consensus"
20701..20809
repeat_region
/note="LTR1 repeat: matches 682..785 of consensus"
20877..21061
repeat_region
/note="LTR1 repeat: matches 1..195 of consensus"
21186..21636
repeat_region
/note="LTR29 repeat: matches 3..498 of consensus"
21609..21737
repeat_region
/note="MER34 repeat: matches 409..543 of consensus"
21711..22076
repeat_region
/note="AluY repeat: matches 2..308 of consensus"
22878..22974
repeat_region
/note="MIR repeat: matches 92..191 of consensus"
23452..23582
repeat_region
/note="FLAM.C repeat: matches 1..131 of consensus"
23583..23624
repeat_region
/note="21 copies 2 mer ta 76% conserved"
23735..24041
repeat_region
/note="AluSg1 repeat: matches 1..308 of consensus"
24045..24072
repeat_region
/note="14 copies 2 mer tt 89% conserved"
25274..25382
repeat_region
/note="AluYb repeat: matches 1..289 of consensus"
25583..25747
misc_feature
/note="match: GSS B63622"

```

```
repeat_region 25959..26012
/Note="HERVFB21 repeat: matches 23..79 of consensus"
repeat_region 26144..26456
/Note="AluY repeat: matches 1..311 of consensus"
repeat_region 27281..27432
/Note="AluJb repeat: matches 137..289 of consensus"
repeat_region 27436..27507
/Note="L1M4 repeat: matches 5360..5434 of consensus"
repeat_region 27536..27600
/Note="L1M4 repeat: matches 2220..2283 of consensus"
repeat_region 27601..27975
/Note="L1M1A1 repeat: matches 3..365 of consensus"
repeat_region 27976..28273

alignment_scores:
  Quality: 372.00      Length: 120
  Ratio: 3.758        Gaps: 2
  Percent Similarity: 82.500  Percent Identity: 60.000

alignment_block:
US-09-327-750D-34 x HS635G19 ..
Align seg 1/1 to: HS635G19 from: 1 to: 69648

1 MetAlaSerLysLysValLysGlnValLeuAspLeuThrValGluLysAs 17
||||| ||||| : : : : : ||||| : : : : :
64293 ATGGAGTCCAAAGAGAGACTAGCGGCAACAACTCAACGGGNAATGC 64342

17 pLysLysAsnLysLysGlyGly...LysAlaSerLysGlnSerGluGluG 33
: : : : : : : : : : : : : : : : : : : : :
64343 CCAACAAGAAACCAAGAGGGAGGAGCGGCCCCACGACGAATCAAGAAG 64392

33 LuSerHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnVal 49
||||| ||||| : : : : : ||||| : : : : :
64393 AATCCCGCCATTTTGGAGGGGGTGAAGGCCAGAGCCTGGAGGAATATC 64442

50 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 65
||||| : : : : : ||||| : : : : : ||||| : : : : :
64443 AGCGGGGGCGAGCTAGCGCACTGTCCCTAATTTTCGATGGGCCATACC 64492

65 oAsnArgHisValAspHisSerGlyGlyGluGluValGlyArgPheV 82
||||| ||||| : : : : : ||||| : : : : : ||||| : : : : :
64493 TAATAGGCATATTGAGCACAATGAAGCGAGAGATGATGTAGAAAGTTTG 64542

82 alGlyGlnValMetGluAlaLysArgHisSerLysGluGlnGlnMetArg 98
||||| ||||| : : : : : ||||| : : : : : ||||| : : : : :
64543 TAGGGCAGATGATGGAAATCAAGAGAAAGACTAGGGGAACAGCAGATGAGG 64592

99 ProTyrThrArgPheArgThrProGluProAspAsnHistyrAspPheCy 115
||| ||||| : : : : : ||||| : : : : : ||||| : : : : :
64593 CACTATATGCGTCTCCAACTCTGAACCTGACACCACTTATGACTTTTG 64642

115 sLeullePro 118
||||| |||||
64643 CCTCATACCT 64652

seq_name: gb_pr.HUMOGC

seq_documentation_block:
LOCUS HUMOGC 891 bp mRNA PRI 07-MAR-1995
DEFINITION Human unknown protein from clone pHGR74 mRNA, complete cds.
ACCESSION M38188 X56942
VERSION M38188.1 GI:189378
KEYWORDS
SOURCE Human ovarian granulosa cell line, cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 891)
AUTHORS Rapp,G., Freudenstein,J., Klaudiny,J., Mucha,J., Wempe,F.,
Zimmer,M. and Scheit,K.H.
TITLE Characterization of three abundant mRNAs from human ovarian
granulosa cells
```

```
JOURNAL DNA Cell Biol. 9 (7), 479-485 (1990)
MEDLINE 91025550
COMMENT Draft entry and computer-readable sequence for [DNA 9, 479-485
(1990)] kindly submitted
by K.H.Scheit, 27-AUG-1990.
FEATURES
  Location/Qualifiers
    Source
      1..891
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="pHGR74"
        /cell_type="granulosa"
        /tissue_type="ovary"
        <1..891
        /note="protein of unknown function"
    CDS
      312..647
        /note="protein of unknown function"
        /codon_start=1
        /protein_id="AAA63232.1"
        /db_xref="GI:189379"
        /translation="MANIHOENEMEOPMONGEEDRPLGGEGEGHQAGNRRRQARRLA
        PNRWAIPIRQINDGMGGDDMEIFMEEMREIRKKRELQRLNCLRLMGLSNNHHD
        HHDEFCLMP"
        361..534
        /note="protein of unknown function"
        /codon_start=1
        /protein_id="AAA63233.1"
        /db_xref="GI:189380"
        /translation="MERKALWEEVKATSLQEIDGDLRLDPLIFDGPYPITGRSMGW
        VEMEMINKYSWRR"
BASE COUNT 251 a 182 c 224 g 234 t
ORIGIN

alignment_scores:
  Quality: 171.50      Length: 126
  Ratio: 2.257        Gaps: 3
  Percent Similarity: 60.317  Percent Identity: 32.540

alignment_block:
US-09-327-750D-34 x HUMOGC ..
Align seg 1/1 to: HUMOGC from: 1 to: 891

12 LeuThrValGluLysAspLysLysAsnLysLysGlyGlyLysAlaSerLy 28
||||| ||||| : : : : : ||||| : : : : :
267 TTACTTGTGTTCAAAGAAACCAACCAACCAACCAACCAACCAACCAACCA 316

28 sGlnSerGluGluSerHisHisLeuGluGluVal..... 40
: : : : : : : : : : : : : : : : : : : : :
317 AAATATTCACCAGAAACCAACCAACCAACCAACCAACCAACCAACCAAT 366

41 .....GluAsnLysLysProGlyGlyAsn 48
||||| : : : : : ||||| : : : : : ||||| : : : : :
367 AGGAGACGCCCTTTGGGAGGAGGTGAAGGCCACCAGCCTGCAGGAAAT 416

49 ValArgArgLysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 65
||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
417 CGACGGGAGCAGCTCGCCGACTTGCCTTAATTTTCGATGGGCCATACC 466

65 oAsnArgHisValAspHisSerGluGly.....GlyGluGluValGlyA 80
||||| : : : : : ||||| : : : : : ||||| : : : : :
467 CAATAGCAGATCAATGATGGGATGGGTGGAGATGATGATGATGATGATAA 516

80 rgPheValGlyGlnValMetGluAlaLysArgHisSerLysGluGlnGln 96
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
517 TATTATGGAGGAGATGAGAGAAATCAGAAGAAACCTTAGGGAGCTGCAG 566

97 MetArgProTyrThrArgPheArgThrProGluProAspAsnHistyrAs 113
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
567 TTGAGGAATGTGCTGCTATCCTTATGGGGGAGCTCTCTTAATCACCATGA 616

113 p.....PheCysLeuIlePro 118
||||| |||||
```



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

```
/crone_uid= nhl_mcc_10
/lab_host="DH10B-R"
/note="Vector: pOTB7"
```

**JOURNAL** Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA

**FEATURES**

<b>SOURCE</b>	Location/Qualifiers 1..519 /organism="Rattus norvegicus" /strain="Sprague-Dawley" /db_xref="taxon:10116" 1..519 /gene="Nade" 44..406 /gene="Nade"
<b>gene</b>	/note="Involved in the common neurotrophin receptor p75NTR-mediated signal transduction; NADe" /codon_start=1 /product="p75NTR-associated cell death executor" /protein_id="AAF75130.1" /db_xref="GI:8452896"
<b>CDS</b>	/translation="MEQLNGQEDRPVGGEGHQPAANNHHNNHNNHHNRHRRGARRLAPNFRAIPNQMDGLGGDDMEMFEIRRKRLRELQLRNLRLILMGELSLHP"

**BASE COUNT** 225 a 170 c 205 g 193 t

**ORIGIN**

```

/codon_start=1
/product="p75NTR-associated cell death executor; ovarian granulosa cell protein (13kd)"
/protein_id="AAH03190.1"
/db_xref="GI:13112031"
/translation="MANIHQENEMEQPMQNBEDRPLGGGEGHQPNAGRRRQARRLA
PNFRWAIPIQNDMGDGDMMFMEIRRKRLRELQLRNLRLILMGELSLLHHD
HHDEFCLMP"

```

**alignment\_scores:**  
Quality: 170.00 Length: 96  
Ratio: 2.656 Gaps: 2  
Percent Similarity: 66.667 Percent Identity: 39.583

**alignment\_block:**  
US-09-327-750D-34 x BC003190 ..

Align seg 1/1 to: BC003190 from: 1 to: 793

```

29 GlnSerGluGluGluSerHisHisLeuGluGluValGluAsnLysLysPr 45
||||| |||||..... ||| ||||: ||||: ||||:
241 CAGAATGGAGAGGAAGACCCCTTTGGGAGGAGGTGAAGGCCACCAGCC 290
45 oGlyGlyAsnValArgLysValArgLysValProAsnPheLeuT 62
||||| ||| ..... ||||| ||||| ||||| |||||
291 TGCAGGAATCGACGGGACAGGCTCGCCACTTCGCCCTTAATTTTCGAT 340
62 rpAlaIleProAsnArHisValAspHisSerGluGly..... GlyGlu 76
||||| ||||| ||||| ..... ||| ||||: ||||:
341 GGGCATACCAATAGGCAGATCAATGCGATGGTGATGGAGATGGAGAT 390
77 GluValGlyArgPheValGlyGlnValMetGluAlaLysArgHisSerLy 93
::::: ||||: ||||: ||||: ||||: ||||: ||||:
391 GATATGGAATATTTCATGGAGGAGATCAGAGAAATCAGAGAAAACCTTAG 440
93 sGluGlnGlnMetArgProThrArgPheArgThrProGluProAspA 110
||| ||||: ||||: ||||: ||||: ||||: ||||: ||||:
441 GGAGCTCGAGTGGAGGAATTCGCTATCTCTTATGGGGGAGCTCTCTA 490
110 snHisTyraSP.....PheCysLeuIlePro 118
||||: ||||: ||||: ||||: ||||: ||||: ||||:
491 ATCACCATGACCATCATGATGAATTTTCCTTATGGCT 528
seq_name: gb_ro-AF187065

```

**seq\_documentation\_block:**  
LOCUS AF187065 519 bp mRNA ROD 11-JUN-2000  
DEFINITION Rattus norvegicus p75NTR-associated cell death executor (Nade)  
mRNA, complete cds.  
ACCESSION AF187065  
VERSION AF187065.1 GI:8452895  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

**REFERENCE**  
1 (bases 1 to 519)  
AUTHORS Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,  
Suwanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.  
TITLE NADe, a p75NTR-associated cell death executor, is involved in  
signal transduction mediated by the common neurotrophin receptor  
p75NTR  
J. Biol. Chem. 275 (23), 17566-17570 (2000)

**MEDLINE**  
20298829  
**REFERENCE**  
2 (bases 1 to 519)  
AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suwanto,P.,  
Hanaoka,T., Li,Y., Irie,S. and Sato,T.  
TITLE Direct Submission

**BASE COUNT** 142 a 130 c 136 g 111 t

**ORIGIN**

```

22 LysGlyGlyLysAlaSerLysGlnSerGluGluGluSerHisHisLeuGl 38
||||| ||: ||: |||||..... ||||: ||||: ||||:
95 GAGGGCCACCAGCCTCTGCTCAAACAACAACAACCAACCAACCAACAA 144
38 uGluValGluAsnLysLysProGlyGlyAsnValArgArgLysValArg 55
::::: ||||| ||||: ||||: ||||: ||||: ||||: ||||:
145 CCACACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 194
55 rgLeuValProAsnPheLeuTrpAlaIleProAsnArghisValAsphIs 71
||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
195 GACTTGCCCTAAC TTTCGATGGCCATTCGCCACACGACGATGAATCAT 244
72 SerGluGly.....GlyGluGluValGlyArgPheValGlyGlnValMe 86
::: ||| |||||..... ||||: ||||: ||||: ||||:
245 GGGTTGGTGGAGATGGAGATGATATGGAATGTTTCATGGAGGAGATGAG 294
86 tGluAlaLysArgHisSerLysGluGlnGlnMetArgProThrArgp 103
||| ||||| ||||: ||||: ||||: ||||: ||||: ||||:
295 AGAGATCCCGAGAGAGCTTTAGGGAGCTGCAGTTGAGAAATTTGTCGCTA 344
103 heArgThrProGluProAspAsnHisTyraSP.....PheCys 115
:: ||| ||| ..... ||||| ||||: ||||: ||||:
345 TTCCTATGGGGAGCTCTCTATATCACCAGCACCATCAGATGAATTCCTGC 394
116 LeuIlePro 118
||||: ||||
395 CTATGCGCT 403
seq_name: gb_sts:G72708

```

**seq\_documentation\_block:**  
LOCUS G72708 421 bp DNA STS  
DEFINITION MARC 4953-4954:991939031:1 SCF - porcine spleen Sus scrofa STS  
genomic, sequence tagged site.  
ACCESSION G72708  
VERSION G72708.1 GI:15146738  
KEYWORDS STS.  
SOURCE pig.  
ORGANISM Sus scrofa



```

||||:||||
162 AACAAAATCTCATCGGCAATGTCCACGAGAAACCAAGAGATGGA 211
34 rHisHisLeuGluValGluAsnLysLysProGlyGly..... 47
212 GCAGCCCTGCAGAAATGGACAGGAAGACCCCTGTGGGAGGAGGTGAGG 261
47 .....
262 GCCCAGCCCTGCTGCAAAACAACAACAACAACAACAACCAACCAACCAAC 311
48 .....AsnValArgArgLysValArgArgLeuValProAsnPheLeuTr 62
312 AACCAACCAAGAGAGCGCGCTGCGGACTTGCCTTAACCTCCGATG 361
62 pAlaIleProAsnArgHisValAspHisSerGluGly.....GlyCluG 77
362 GGCCATTCCCAACGACGAGATGAATGACGGGTTGGGTGGAGATGAGATG 411
77 luValGlyArgPheValGlyGlnValMetGluAlaLysArgHisSerLys 93
412 ATATGGAATGTTCTACGACGAGATGAGAGATCCGGAGAAAGCTTAGG 461
94 GluGlnGlnMetArgProTyrThrArgPheArgThrProGluProAspAs 110
462 GAGCTACAGCTGAGAAATGTTCTACGATCCTTATGGGGGAGCTGTCTAA 511
110 nHisTyrAsp.....PheCysLeuIlePro 118
512 CCACGAGATCACCAGATGATGAATTCGCCTTATGCCT 548

seq_name: gb_ro:AF097440

seq_documentation_block:
LOCUS AF097440 854 bp mRNA ROD 13-APR-1999
DEFINITION Mus musculus brain expressed x-linked protein 3 (Bex3) mRNA,
complete cds.
ACCESSION AF097440
VERSION AF097440.1 GI:4580593
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 854)
TITLE Brown,A.L. and Kay,G.F.
JOURNAL Bex1, a gene with increased expression in parthenogenetic embryos,
MEDLINE Hum. Mol. Genet. 8 (4), 611-619 (1999)
REMARK 99172070
Erratum: [[published erratum appears in Hum Mol Genet 1999
May;8(5):943]]
REFERENCE 2 (bases 1 to 854)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
MEDLINE Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
source
1..854
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/db_xref="dbEST:AA272375"
/chromosome="X"
/map="near Plp"
/tissue_type="pooled organs"
1..854
/gene="Bex3"
172..546
/gene="Bex3"
/codon_start=1
/product="brain expressed x-linked protein 3"
/protein_id="AAD24431.1"

```

```

/db_xref="GI:4580594"
/translation="MANVHQNEMEOPLONGDRPVGGEHQPAAANNHNNHNNH
HNHRRGOARLAPNFWAIPNROMDCLGDDMEMFMEIRKRLRELQLRNC
LRIELGELSNHHHDEFCLMP"
BASE COUNT 237 a 212 c 228 g 177 t
ORIGIN

alignment_scores:
Quality: 145.00 Length: 129
Ratio: 2.014 Gaps: 3
Percent Similarity: 55.814 Percent Identity: 28.682
alignment_block:
US-09-327-750D-34 x AF097440 ..
Align seg 1/1 to: AF097440 from: 1 to: 854
18 LysLysAsnLysLysGlyLysAlaSerLysGlnSerGluGluGlue 34
157 AACAAATCTCATCGGCAATGTCCACGAGAAACCAAGAGATGGA 206
34 rHisHisLeuGluValGluAsnLysLysProGlyGly..... 47
207 GCAGCCCTGCAGAAATGGACAGGAAGACCCCTGTGGGAGGAGGTGAGG 256
47 .....
257 GCCCAGCCCTGCTGCAAAACAACAACAACAACAACAACCAACCAACCAAC 306
48 .....AsnValArgArgLysValArgArgLeuValProAsnPheLeuTr 62
307 AACCAACCAAGAGAGCGCGCTGCGGACTTGCCTTAACCTCCGATG 356
62 pAlaIleProAsnArgHisValAspHisSerGluGly.....GlyCluG 77
357 GGCCATTCCCAACGACGAGATGAATGACGGGTTGGGTGGAGATGAGATG 406
77 luValGlyArgPheValGlyGlnValMetGluAlaLysArgHisSerLys 93
407 ATATGGAATGTTCTACGACGAGATGAGAGATCCGGAGAAAGCTTAGG 456
94 GluGlnGlnMetArgProTyrThrArgPheArgThrProGluProAspAs 110
457 GAGCTACAGCTGAGAAATGTTCTACGATCCTTATGGGGGAGCTGTCTAA 506
110 nHisTyrAsp.....PheCysLeuIlePro 118
507 CCACGAGATCACCAGATGATGAATTCGCCTTATGCCT 543

seq_name: gb_pr:HSV351F8

seq_documentation_block:
LOCUS HSV351F8 45678 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from cosmid V351F8, between markers DXS366 and
DXS87 on chromosome X contains ESTs.
ACCESSION 270719
VERSION 270719.1 GI:1261915
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 (bases 1 to 45678)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Whiteley,M.
JOURNAL Direct Submission
COMMENT Submitted (09-APR-1995) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk
IMPORTANT: This sequence is the entire insert of clone V351F8. The
true left end of clone V351F8 is at 1 in this sequence. The true
right end of clone V351F8 is at 45678.
V351F8 is from the human chromosome X-specific cosmid library.
FEATURES
Location/Qualifiers

```

```
source
1. .45678
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="GHC-351F8"
/clone_lib="SCcv"
1. .70
repeat_region
/partial
/note="Alu repeat: matches 80. .1 of consensus"
642. .930
repeat_region
/note="Alu repeat: matches 1. .308 of consensus"
1406. .5179
repeat_region
/note="L1 element fragment"
5243. .5276
repeat_region
/note="17 copies of 2 mer 82 % conserved"
5279. .5567
repeat_region
/partial
/note="Alu repeat: matches 304. .1 of consensus"
9975. .10369
repeat_region
/note="MSTA element fragment"
10399. .12017
repeat_region
/note="MSTAR element fragment"
11318. .11978
repeat_region
/note="THEIBR element fragment"
11715. .11846
repeat_region
/note="MLTIR element fragment"
12074. .12170
repeat_region
/note="MSTA element fragment"
12227. .12486
repeat_region
/note="MSTA element fragment"
12437. .12502
repeat_region
/note="MSTC element fragment"
12756. .13059
repeat_region
/partial
/note="Alu repeat: matches 308. .1 of consensus"
14478. .14650
repeat_region
/note="THEIB element fragment"
14491. .14536
repeat_region
/note="MSTA element fragment"
14715. .14831
repeat_region
/note="THEIB element fragment"
15099. .15533
repeat_region
/note="L1 element fragment"
15534. .15626
repeat_region
/note="31 copies of 3 mer 85 % conserved"
15566. .15619
repeat_region
/note="3 copies of 18 mer 98 % conserved"
16352. .16436
repeat_region
/note="L1 element fragment"
16728. .16763
repeat_region
/note="9 copies of 4 mer 94 % conserved"
16764. .17054
repeat_region
/partial
/note="Alu repeat: matches 308. .1 of consensus"
19420. .19449
repeat_region
/note="15 copies of 2 mer 87 % conserved"
complement(19537. .20013)
misc_feature
/note="match: 3' EST N51315 clone 283089"
20071. .20114
repeat_region
/note="22 copies of 2 mer 98 % conserved"
20073. .20112
repeat_region
/note="10 copies of 4 mer 100 % conserved"
20073. .20126
repeat_region
/note="3 copies of 18 mer 87 % conserved"
23088. .23300
repeat_region
/note="L1 element fragment"
23311. .23490
repeat_region
/note="MLTIA element fragment"
23318. .23489
repeat_region
/note="MLTIB element fragment"
23362. .23491
repeat_region
/note="MLTIC element fragment"

repeat_region
23507. .24407
/note="L1 element fragment"
24376. .24462
repeat_region
/note="MSTC element fragment"
24379. .24491
repeat_region
/note="MSTA element fragment"
24688. .24738
repeat_region
/note="MSTA element fragment"
24810. .25016
repeat_region
/note="L1 element fragment"
25254. .25421
repeat_region
/note="L1 element fragment"
25439. .25696
repeat_region
/note="MER25 element fragment"
26284. .26343
repeat_region
/note="MLTID element fragment"
26428. .26491
repeat_region
/note="MLTID element fragment"
26465. .26518
repeat_region
/note="MLTIE element fragment"
27446. .27667
repeat_region
/note="L1 element fragment"
28954. .29247
repeat_region
/partial
/note="Alu repeat: matches 308. .1 of consensus"
30160. .30325
repeat_region
/note="2 copies of 83 mer 98 % conserved"
32588. .32875
misc_feature
/note="match: 5' EST H68599 clone 239077"
complement(32825. .33230)
misc_feature
/note="match: 3' EST H68239 clone 289077"
35551. .35802
repeat_region
/partial
/note="Alu repeat: matches 1. .260 of consensus"
35804. .35839
repeat_region
/note="18 copies of 2 mer 83 % conserved"
38625. .38705
repeat_region
/note="MLTIR element fragment"
38747. .38839
repeat_region
/note="MLTIR element fragment"
38949. .39032
repeat_region
/note="MLTIE element fragment"
38965. .39076
repeat_region
/note="MLTID element fragment"
BASE COUNT 14260 a 9135 c 9327 g 12956 t
ORIGIN

alignment_scores:
Quality: 138.00 Length: 127
Ratio: 2.000 Gaps: 3
Percent Similarity: 54.331 Percent Identity: 29.921

alignment_block:
US-09-327-750D-34 x HSV351F8 ..
Align seg 1/1 to: HSV351F8 from: 1 to: 45678

12 LeuThrValGluLysAspLysLysAsnLysGlyGlyLysAlaSerLy 28
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
19408 TTATCTTCTAGGAAGAAAAAAGAAAAAAGAAAAAAGAAAAAATCAACAT 19457

28 sGlnSer.....GlUG 32
||||:
19458 GGAAATGTCCCAAGGAAAAACAAGTTGTGGAGAGGCCCAAGTCGAGA 19507

32 luGluSerHisHisLeuGluValGluAsnLysLysProGlyGlyAsn 48
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
19508 ATGAAGCCCGCGCTTTAGGAGGTGGTGAATACAGGAGCCTGGAGGAAT 19557

49 ValArgArgLysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 65
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
19558 GTTAAAGGGTTTGGGGCTCCACCTGCCCGGGGTTTGGAGAGGATGTGCC 19607
```

STS	44.	.149
primer_bind	44.	.67

139. 528

gene

CDS







OM of: US-09-327-750D-33 to: EST:\* out\_format : pfs

Date: Mar 11, 2002 2:16 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101153\_20308/app\_query.fasta\_1.1472  
-DB=EST -OPMT=fastap -SUFFIX=tst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09327750\_ECGNL\_1\_5654  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-33

Query length: 128

Database: EST:\*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 4085.940000

score\_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
gb_est1:AL022932	+	646.00	1047.10	3.7e-49	551	! AL022932 v8130b28 Beddington mc
gb_est1:AW536404	-	646.00	1045.88	4.3e-49	642	! AW536404 G0104A11-3 NIA Mouse E
gb_est2:BG004920	+	643.00	1040.25	8.9e-48	707	! BG004920 H3025D01-3 NIA Mouse I
gb_est2:BG277659	+	633.00	1025.69	5.7e-48	575	! BG277659 ux47c11.y1 Soares.NMMA
gb_est1:BG870503	+	633.00	1023.51	7.6e-48	756	! BG870503 602791422F1 NCI CGAP.S
gb_est1:BE654459	+	615.00	997.65	2.1e-46	499	! BE654459 UI-M-AJ1-aha-f-10-0-UI
gb_est1:AW536974	+	599.00	970.56	6.8e-45	577	! AW536974 G0110H10-3 NIA Mouse E
gb_est2:BF607762	+	554.00	895.17	1.1e-40	785	! BF607762 MYL_000704 Mouse 9-day
gb_est2:BG228077	-	553.00	895.87	9.8e-41	587	! BG228077 ux47c11.x1 Soares.NMMA
gb_est2:BF608209	+	516.00	835.09	2.4e-37	650	! BF608209 MYL_001088 Mouse 9-day
gb_est1:BE291071	+	505.50	781.94	1.9e-36	583	! BE291071 G01086311F1 NCI CGAP.M
gb_gss:AZ936393	-	493.50	798.78	2.5e-35	637	! AZ936393 2M0193L05F Mouse 10kb
gb_est2:BG669326	+	489.50	795.78	3.7e-35	412	! BG669326 DRNAEB11 Rat DRG Libr
gb_est2:BL183527	-	480.50	776.95	4.1e-34	701	! BL183527 UNL-P-FN-by-f-07-0-UNL
gb_est1:AW536634	-	472.00	765.12	1.9e-33	549	! AW536634 G0106H08-3 NIA Mouse E
gb_est2:BG655001	+	471.50	764.34	2.1e-33	547	! BG655001 ib44f06.y1 HR85 islet
gb_est2:BG472167	+	471.50	763.17	2.4e-33	634	! BG472167 602513894F1 NIH_MGC_16
gb_est2:BF971303	+	471.50	762.87	2.5e-33	658	! BF971303 602273150F1 NIH_MGC_84
gb_est2:BG714823	+	471.50	762.68	2.6e-33	674	! BG714823 602677146F1 NIH_MGC_96
gb_est1:AV702285	+	471.50	762.49	2.6e-33	690	! AV702285 AV702285 ADB Homo sapi
gb_est1:AV705808	+	471.50	762.48	2.6e-33	691	! AV705808 AV703808 ADB Homo sapi
gb_est1:AL526209	-	471.50	762.39	2.7e-33	695	! AL526209 AL526209 LTI_NFL003_NE
gb_est1:BG820179	+	471.50	762.32	2.7e-33	699	! BG820179 602782222F1 NCI CGAP.F
gb_est1:BE790774	+	471.50	761.48	3.0e-33	783	! BE790774 G01581956F1 NIH_MGC_7
gb_est1:AL526247	+	471.50	761.37	3.0e-33	794	! AL526247 AL526247 LTI_NFL003_NE
gb_est1:AV702643	+	468.50	757.71	4.8e-33	683	! AV702643 AV702643 ADB Homo sapi
gb_est1:BE314909	+	468.00	756.01	4.7e-33	594	! BE314909 G01140074F1 NIH_MGC_9
gb_est1:AL520931	-	468.00	756.51	5.7e-33	717	! AL520931 AL520931 LTI_NFL004_NE
gb_est2:BE889313	+	468.00	755.92	6.1e-33	772	! BE889313 G01513201F1 NIH_MGC_71
gb_est2:BG709503	+	468.00	755.54	6.4e-33	810	! BG709503 602674781F1 NIH_MGC_96
gb_est2:BG707398	+	468.00	755.29	6.6e-33	835	! BG707398 602672810F1 NIH_MGC_96
gb_est2:BG285825	+	468.00	754.30	7.5e-33	946	! BG285825 602360917F1 NIH_MGC_93
gb_est2:BF038787	+	468.00	753.53	8.3e-33	1042	! BF038787 G01462119F1 NIH_MGC_6
gb_est1:AI929703	+	464.50	752.59	9.4e-33	576	! AI929703 au63f04.y1 Schneider F
gb_est2:BG715659	+	464.00	750.65	1.2e-32	734	! BG715659 602676924F1 NIH_MGC_96
gb_est2:BG714974	+	461.00	745.07	2.5e-32	726	! BG714974 602677263F1 NIH_MGC_96
gb_est2:BF967675	-	461.00	744.36	2.7e-32	793	! BF967675 602287335F1 NIH_MGC_96
gb_est2:BG707734	+	461.00	744.12	2.8e-32	817	! BG707734 602671235F1 NIH_MGC_96
gb_est2:BG705843	+	457.00	737.68	6.3e-32	813	! BG705843 602669329F1 NIH_MGC_96
gb_est2:BG709325	+	456.50	736.72	7.2e-32	828	! BG709325 602673345F1 NIH_MGC_96

gb\_est1:AI291270 + 455.00 738.14 6.0e-32 511 ! AI291270 qml6e06.x1 NCI\_CGAP  
gb\_est1:AL024066 + 453.00 734.67 9.3e-32 526 ! AL024066 p8225a45 Beddington  
gb\_est1:BE266012 + 453.00 733.49 1.1e-31 610 ! BE266012 601194731F1 NIH\_MGC  
gb\_est1:AI291126 + 453.00 732.75 1.2e-31 669 ! AI291126 qml5f02.x1 NCI\_CGAP  
gb\_est2:BF237433 + 453.00 732.27 1.3e-31 711 ! BF237433 601842108F1 NIH\_MGC  
gb\_name: gb\_est1:AL022932

seq\_documentation\_block:  
LOCUS AL022932 551 bp mRNA EST 29-DEC-1999  
DEFINITION v8130b28 Beddington mouse dissected endoderm Mus musculus CDNA  
clone 528-3B10 5', mRNA sequence.

ACCESSION AL022932  
VERSION AL022932.1 GI:6645505  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 551)  
AUTHORS Harrison,S.M., Dunwoodie,S.L., Arkell,R.M., Lehrach,H. and Beddington,R.S.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissue constituents of the gastrulating mouse embryo  
JOURNAL Development 121 (8), 2479-2489 (1995)  
MEDLINE 95401865

COMMENT Contact: Wiles,M., Lehrach,H. and Ayner,P.  
EEC Mouse Transcript Mapping Consortium  
Genoscope - CNS  
Email: pavner@pasteur.fr

Clone available from Ressourcenzentrum, Heubnerweg 6, D-14059 Berlin, Germany. Web site http://www.rzpd.de  
Seq primer: CCGTCCGAATCCGGGT;  
High quality sequence only submitted.  
Vector: pSport1; site\_1: NotI; site\_2: SalI;  
Cloned unidirectionally.

Dissected endoderm 7.5 days.  
Average insert size: 1.2 kb (range: 0.2 - 2.kb).  
Location/Qualifiers  
1..551  
/organism="Mus musculus"  
/strain="C57Bl6 x DBA"  
/db\_xref="taxon:10090"  
/clone="528-3B10"  
/clone\_lib="Beddington mouse dissected endoderm"  
/tissue\_type="dissected endoderm"  
/dev\_stage="7.5 dpc"

FEATURES

source

/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally ~ High quality sequence only submitted. ~ Average insert size: 1.2 kb (range: 0.2 - 2.kb)"  
BASE COUNT 153 a 124 c 191 g 83 t  
ORIGIN

alignment\_scores:  
Quality: 646.00 Length: 128  
Ratio: 5.168 Gaps: 0  
Percent Similarity: 97.656 Percent Identity: 92.969

alignment\_block:  
US-09-327-750D-33 x AL022932 ..

Align seg 1/1 to: AL022932 from: 1 to: 551

1 MetGlyLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspH 17  
|||||

142 ATGGAGTCCAAGATCAAGCGCTGAAATATCTCAACATGGAGATGACCA 191  
|||||

17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleLysArgG 34  
|||||

192 TCAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 241  
|||||

```

34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProAtrg 50
|||||:||||| ||| |||||||
242 AGCCAGCTGGGCCCTCGACTCCGAGGCTGGCAAAACATGTGGCGCTAGA 291
|||||:||||| ||| |||||||
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyrArgTr 67
|||||:||||| ||| |||||||
292 GGAGGTGCGAGGGGTTCGGGTTCGGCAGCCCATCGCTCACTATAGATG 341
|||||:||||| ||| |||||||
67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGluA 84
|||||:||||| ||| |||||||
342 GGACCTGATCGACAGGGTTCGGGAGCCCGCCAGGGAAGGATGAGAGGAGA 391
|||||:||||| ||| |||||||
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
|||||:||||| ||| |||||||
392 ACGTACAGAGGTTCGGGGTGTGTGAGACAGCTCATGGAGAAGCTGAGG 441
|||||:||||| ||| |||||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||:||||| ||| |||||||
442 GAAGGCAGCTGAGCCACAGCCTGCGGGCGGTAGCACTGACCGGCTCA 491
|||||:||||| ||| |||||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:||||| ||| |||||||
492 TCATGACCACCATGATGAGTTTTCGCTCATGCC 525
|||||:||||| ||| |||||||

```

seq\_name: gb\_est1:AW536404

```

seq_documentation_block:
LOCUS AW536404 642 bp mRNA EST 31-AUG-2000
DEFINITION G0104A11-3 NTA Mouse E7.5 Embryonic Portion cDNA Library Mus
musculus cDNA clone G0104A11 3', mRNA sequence.
ACCESSION AW536404
VERSION AW536404.1 GI:7178821
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
20381348
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: G0104 row: A column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 642
POLYA=Yes.

```

```

FEATURES
    source
        1..642
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="G0104A11"
            /clone_lib="NTA Mouse E7.5 Embryonic Portion cDNA Library"
            /sex="unknown"
            /dev_stage="7.5dpc Embryo"
            /lab_host="DH10B"
            /note="Vector: pSPOR1 (Gibco/BRL Life Technology);
            Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
            6 Embryo. The double-stranded cDNA was synthesized by
            Gibco's kit with an Oligo(dT) primer [NotI primer-adapter
            from GibcoBRL]
            [5'-pGACTAGTCTTATAGATCGAGCGCGCCCTTTTTCCTTTTTCCTTTT-3']
            from 0.51ug of mRNA . The double-stranded cDNAs were
            treated with T4 DNA polymerase and purified by

```

ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPOR1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang and Minoru S. H. Ko."

BASE COUNT 145 a 182 c 133 g 182 t  
ORIGIN

alignment\_scores:  
Quality: 646.00 Length: 128  
Ratio: 5.168 Gaps: 0  
Percent Similarity: 97.656 Percent Identity: 92.969

alignment\_block:  
US-09-327-750D-33 x AW536404/rev ..

Align seg 1/1 to reverse of: AW536404 from: 1 to: 642

```

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHi 17
|||||:||||| ||| |||||||
606 ATGGAGTCCAAAGATCAAGGCGTGAATAATCTCAACATGGAGAATGACCA 557
|||||:||||| ||| |||||||
17 sGlnLysLysGluGluLysGluGlnAspThrIleLysArgG 34
|||||:||||| ||| |||||||
556 TCAGAAAAAGAGGAGAGGAAGAAAAACCAAGATACCATCAGAAAGG 507
|||||:||||| ||| |||||||
34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProAtrg 50
|||||:||||| ||| |||||||
506 ACCGAGCTGTGGCCCTCGACTCCGAGGCTGGCAAAACATGTGCACCTAGA 457
|||||:||||| ||| |||||||
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyrArgTr 67
|||||:||||| ||| |||||||
456 GCAGGTGCGCAGGCGGTTCGGGTTCGGCAGCCCATCGCTCACTATAGATG 407
|||||:||||| ||| |||||||
67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGluA 84
|||||:||||| ||| |||||||
406 GACCTGATCGACAGGGTTCGGGAGCCCGCCAGGATGAGAGAGGAGA 357
|||||:||||| ||| |||||||
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
|||||:||||| ||| |||||||
356 ACGTACAGAGGTTCGGGGTGTGTGAGACAGCTCATGGAGAAGCTGAGG 307
|||||:||||| ||| |||||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||:||||| ||| |||||||
306 GAAAGGCAGCTGAGCCACAGCCTGCGGGCGGTAGCACTGACCGGCTCA 257
|||||:||||| ||| |||||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:||||| ||| |||||||
256 TCATGACCACCATGATGAGTTTTCGCTCATGCC 223
|||||:||||| ||| |||||||

```

seq\_name: gb\_est2:BG064920

```

seq_documentation_block:
LOCUS BG064920 707 bp mRNA EST 26-JAN-2001
DEFINITION H3025D01-3 NTA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3025D01 3', mRNA sequence.
ACCESSION BG064920
VERSION BG064920.1 GI:12547483
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 707)

```



```
US-09-327-750D-33 x BG277659
Align seg 1/1 to: BG277659 from: 1 to: 575

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHI 17
192 ATGGAGTCCAAAGATCAAGCGGTCAAAAATCTCAACATGGAGAATGACCA 241
17 sGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysArgG 34
242 TCAGAAAAGAGGAGAGGAAGAAAGAAAGCCACCAAGATACCATCAGAAGG 291
34 luProValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
292 AGCCAGCTGTGGCCCTCACCTCCAGCGTGGCAAAACTGTGCACCTAGA 341
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyArgTr 67
342 GGAGGTCCAGAGCGGTTCGGGTTGGCGAGCCCATCGCTCACTATAGATG 391
67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGluA 84
392 GGACCTGATGCAGAGGTTGGGAGCGCCAGGGAAGGATGAGAGAGGAGA 441
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100
442 ACGTACAGAGTTGNGGGTGATGTGAGACAGCTCATGGAGAGCTGAGG 491
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHI 117
492 GAAAGGCAGCTGAGCCACAGCCCTGCGGGCGTTAGCACTACCCCGCTCA 541
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
542 TCATGACCAACCATGATGAGTTGTGCCTCATGCC 575

seq_name: gb_est2:BG870503

seq_documentation_block:
LOCUS BG870503 756 bp mRNA EST 29-MAY-2001
DEFINITION 602791422F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4922644 5',
mRNA sequence.
ACCESSION BG870503
VERSION BG870503.1 GI:14221043
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0842 row: 1 column: 05
High quality sequence stop: 756.
Location/Qualifiers
1..756
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:

NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 207 a 164 c 230 g 155 t
ORIGIN
alignment_scores:
Quality: 633.00 Length: 129
Ratio: 5.064 Gaps: 1
Percent Similarity: 96.899 Percent Identity: 92.248
alignment_block:
US-09-327-750D-33 x BG870503
Align seg 1/1 to: BG870503 from: 1 to: 756
1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHI 17
165 ATGGAGTCCAAAGATCAAGCGGTCAAAAATCTCAACATGGAGAATGACCA 214
17 sGlnLysLysGluGluLysGluGluLysProGlnAspThrIleLysArgG 34
215 TCAGAAAAGAGGAGAGGAAGAAAGAAAGCCACCAAGATACCATCAGAAGG 264
34 luProValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50
265 AGCCAGCTGTGGCCCTCACCTCCAGCGTGGCAAAACTGTGCACCTAGA 314
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyArgTr 67
315 GGAGGTCCAGAGCGGTTCGGGTTGGCGAGCCCATCGCTCACTATAGATG 364
67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGlu 83
365 GGACCTGATGCAGAGGTTGGGAGCGCCAGGGAAGGATGAGAGAGGAG 414
84 AsnValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuAr 100
415 AACCTACAGAGGTTTGGGGTGATGTGAGACAGCTCATGGAGAAGCTGAG 464
100 gGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
465 GGAAGGCAGCTGAGCCACAGCCCTGCGGGCGTTAGCACTACCCCGCTC 514
117 isHisAspHisHisAspGluPheCysLeuMetPro 128
515 ATCATGACCAACCATGATGAGTTTGTGCCTCATGCC 549
seq_name: gb_est1:BE654459
seq_documentation_block:
LOCUS BE654459 499 bp mRNA EST 06-SEP-2000
DEFINITION UI-M-AJ1-aha-f-10-0-UI-r1 NIH_BMAP_MOB_N Mus musculus cDNA clone
UI-M-AJ1-aha-f-10-0-UI 5', mRNA sequence.
ACCESSION BE654459
VERSION BE654459.1 GI:9980372
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 499)
Normalizaton and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 9704477
MEDLINE
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
```

Fax: 301 443 9890

Email: mEst@mail.nih.gov

cDNA Library Preparation: M.B. Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Reverse.

# FEATURES

## source

Location/Qualifiers  
1..499  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-AJ1-aba-f-10-0-UI"  
/clone\_lib="NIH\_BMAP\_MOB\_N"  
/dev\_stage="27-32 days"  
/lab\_host="DHI08 (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The NIH\_BMAP\_MOB\_N library is a normalized library constructed from mouse olfactory bulbs. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."

146 a 108 c 179 g 66 t

BASE COUNT  
ORIGIN

## alignment\_scores:

Quality: 615.00 Length: 123  
Ratio: 5.125 Gaps: 0  
Percent Similarity: 97.561 Percent Identity: 92.683

## alignment\_block:

US-09-327-750D-33 x BE654459 ..

Align seg 1/1 to: BE654459 from: 1 to: 499

1 MetGluSerLysAspGlnGlyAlaLysAsnLeuAsnMetGluAsnAspHi 17  
|||||.....  
131 ATGGAGTCCAAAGATCAAGCGCTGAAATCTCAACATGGAGATGACCA 180  
17 scGlnLysLysGluGluLysGluLysProGlnAspThrIleLysArgG 34  
|||||.....  
181 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 230  
34 luProValValAlaProThrPheGluAlaGlyLysAsnCysAlaProArg 50  
|||||.....  
231 ACCAGCTGTGCGCTGACCTCCGAGGCTGGCAAAACTGTCCACTAGA 280  
51 GlyGlyArgArgPheArgValArgGlnProIleSerHisTyrArgTr 67  
|||||.....  
281 GGAGGTCGAGCGGTTCGGGTTCGGAGCCCATCGCTCACTATAGATG 330  
67 pAspLeuMetHisArgValGlyGluProGlnGlyArgMetArgGluGlu 84  
|||||.....  
331 GGACCTGTAGCAGAGGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 380  
84 snValGlnArgPheGlyGluAspMetArgGlnLeuMetGluLysLeuArg 100  
|||||.....  
381 ACGTACAGAGTTTGGGGGTGATGTGACAGACGCTCATGAGAGCTGAGG 430  
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117  
|||||.....  
431 GAAAGCAGCTGAGGCACACAGCTGCGGGGGTGTAGCACTGACCCGCTCA 480  
117 SHIsAspHisAspGlu 123  
|||||.....  
481 TCATGACCACCATGATGAG 499

seq\_name: gb\_estl:AW536974

seq\_documentation\_block:

LOCUS AW536974 577 bp mRNA EST 31-AUG-2000  
DEFINITION G0110H10-3 NIA Mouse E7.5 Embryonic Portion cDNA Library Mus  
musculus cDNA clone G0110H10 3', mRNA sequence.

ACCESSION AW536974

VERSION AW536974.1 GI:7179391

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 577)  
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac  
M.J., Pantano,S., Sano,Y., Plao,Y., Nagaraja,R., Doi,H., Wood,W.H.  
III, Becker,K.G. and Ko,M.S.H.

TITLE Genome-wide expression profiling of mid-gestation placenta and  
embryo using a 15,000 mouse developmental cDNA microarray  
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)  
20381348

CONTACT: George J. Kargul

LABORATORY OF GENETICS

National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

Plate: G0110 row: H column: 10

Seq primer: -21M13 Forward

High quality sequence stop: 577

POLYA-Yes.

## FEATURES

### source

Location/Qualifiers  
1..577  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="G0110H10"  
/clone\_lib="NIA Mouse E7.5 Embryonic Portion cDNA Library"  
/sex="unknown"  
/dev\_stage="7-5dpc Embryo"  
/lab\_host="DHI08"  
/note="Vector: pSPORT1 (Gibco/BRL Life Technology);  
Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from  
6 Embryo. The double-stranded cDNA was synthesized by  
Gibco's kit with an Oligo(dT) primer (NotI primer-adaptor  
from GibcoBRL)

[5'-pgACTAGTCTAGATCGGAGCGCCCTTTTCTTTTCTTTT-3']  
from 0.5ug of mRNA. The double-stranded cDNAs were  
treated with T4 DNA polymerase and purified by  
ethanol-precipitation. The cDNAs were ligated to  
Lone-linker LL-Sal3 (include SalI sequence). The cDNAs  
were purified by phenol/chloroform and separated from  
free linkers by Centricon 100. Then, cDNAs were amplified  
by long-range high fidelity PCR using Takara's Ex Taq  
polymerase. Then, the cDNAs were purified by  
phenol/chloroform and by Centricon 100. The cDNAs were  
digested with SalI and NotI enzymes. Then, the cDNAs were  
size selected by Gibco's Size Fractionation Column. The  
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid  
vector. The DHI08 E. coli host was transformed with the  
ligation mixture by chemical method. The library was  
constructed by Xiaohong Wang and Minoru S. H. Ko."

137 a 160 c 126 g 154 t

BASE COUNT  
ORIGIN

## alignment\_scores:

Quality: 599.00 Length: 118  
Ratio: 5.209 Gaps: 0  
Percent Similarity: 97.458 Percent Identity: 93.220

## alignment\_block:

US-09-327-750D-33 x AW536974/rev ..





```
alignment_scores:
  Quality: 355.50      Length: 120
  Ratio: 3.665        Gaps: 3
  Percent Similarity: 80.833  Percent Identity: 59.167
```

[illegible]



114 sLeuIlePro 117  
 |||||  
 541 CCTCATACCT 550

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAC85548

seq\_documentation\_block:

ID AAC85548 standard; cDNA; 1364 BP.

XX AAC85548;

XX 04-JUN-2001 (first entry)

XX cDNA encoding CDIFF-4, Incyte ID No. 1990956CB1.

XX Human; cell differentiation; CDIFF; agonist; antagonist; epilepsy;  
 KW cell proliferation; Alzheimer's disease; schizophrenic disorder;  
 KW arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 250..612

FT /\*tag= a

FT /product= "CDIFF-4"

XX WO200119860-A2.

XX 22-MAR-2001.

XX 14-SEP-2000; 2000WO-US25435.

XX 15-SEP-1999; 99US-0154140.

XX 06-DEC-1999; 99US-0169155.

XX (INCY-) INCYTE GENOMICS INC.

XX Tang YT, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;  
 PI Baughn MR, Lu DAM, Bandman O, Shih LL, Patterson C;

XX WPI; 2001-211447/21.

XX P-PSDB; AAB47126.

XX Isolated polypeptides and polynucleotides involved in cell  
 PT differentiation are used for treatment, prevention and diagnosis of  
 PT cell proliferative, developmental and neurological disorders e.g.  
 PT cancer and Alzheimer's disease -

XX Claim 5; Page 121; 137pp; English.

XX The sequences given in AAC85545-72 encode human polypeptides involved  
 CC in cell differentiation (CDIFF). CDIFF polypeptides and agonists of  
 CC these are used to treat a disease or condition associated with  
 CC decreased expression of functional CDIFF. An antagonist of CDIFF is  
 CC used to treat a disease or condition associated with over expression  
 CC of functional CDIFF. CDIFF polypeptides may be used for the treatment,  
 CC prevention and diagnosis of cell proliferative, developmental and  
 CC neurological disorders, such as Alzheimer's disease, schizophrenic  
 CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus  
 CC and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus  
 CC REX-3. This sequence maps to chromosome 1 within the interval from  
 CC 152.2 to 157.4 centiMorgans, to chromosome 3 within the interval from  
 CC 157.4 to 158.0 centiMorgans, and to the X chromosome within the interval  
 CC from 104.9 to 150.3 centiMorgans.

XX Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;

alignment\_scores:

Quality: 355.50 Length: 120

Ratio: 3.665 Gaps: 3

Percent Similarity: 80.833 Percent Identity: 59.167

alignment\_block:

US-09-327-750D-35 x AAC85548

Align seg 1/1 to: AAC85548 from: 1 to: 1364

1 MetAlaSerLysLysGlnValIleLeu...AspLeuThrValGluLysAs 16  
 ||| |||||:|||||: |||||: |||||:  
 250 ATGAGTCCAAAGAGGAACCTAGCGCAACAATCTCAACGGGGAATGC 299  
 16 pLysLysAspLysArgGlyGly...LysAlaSerLysGlnSerGluGluG 32  
 :|||:|||||:|||||: |||||: |||||: |||||:  
 300 CCAACAAGAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 349  
 32 LuProHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnVal 48  
 || |||||:|||||: |||||: |||||: |||||: |||||:  
 350 AATCCGCCCATTTGGGAGGGGTGAGGCCAGAGAGCCTGGAGGAATATC 399  
 49 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 64  
 |||||: |||||: |||||: |||||: |||||: |||||:  
 400 AGCGGGGGCGAGTTAGCGGACTTGTCCTAATTTTCGATGGGCATACC 449  
 64 OAsnArgHisValAspArgAsnGluGlyGluAspValGlyArgPheV 81  
 |||||: |||||: |||||: |||||: |||||: |||||:  
 450 TAATAGGCATATTGAGCACAATCAAGCGAGAGATGATGATAGAAAGTTG 499  
 81 alValGlnGlyThrGluValLysArgLysThrThrGluGlnValArg 97  
 || |||||: |||||: |||||: |||||: |||||: |||||:  
 500 TAGGGCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGACAGATGAGG 549  
 98 ProTyrArgArgPheArgThrProGluProAspAsnHisTyrAspPheCy 114  
 ||| |||||: |||||: |||||: |||||: |||||: |||||:  
 550 CACTATATGCGCTTCCAAACTCTGAACTGACAACTTATGACTTTTG 599  
 114 sLeuIlePro 117  
 |||||  
 600 CCTCATACCT 609

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAC85548

seq\_documentation\_block:

ID AAF23529 standard; DNA; 891 BP.

XX AC AAF23529;

XX XX 22-MAR-2001 (first entry)

XX DE Human NADE DNA.

XX Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;

KW neurogenetic disease; NF-kappaB; ds.

XX Homo sapiens.

XX WO2000075278-A2.

XX PD 14-DEC-2000.

XX PF 07-JUN-2000; 2000WO-US15621.

XX PR 07-JUN-1999; 99US-0327750.

XX PA (UYCO ) UNIV COLUMBIA NEW YORK.

XX Sato T;

XX WPI; 2001-061707/07.

XX New p75-neurotrophin receptor-associated cell death executor (NADE) and  
 PT the gene encoding NADE, useful for modulating the activity of p75NTR  
 PT and for detecting neurodegenerative diseases -  
 XX Disclosure; Fig 1; 134pp; English.









```

PR 10-NOV-1999; 99US-0164647.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
XX
XX WPI; 2001-112727/12.
DR P-PSDB; AAB60474.
XX
XX Human cell cycle and proliferation proteins and polynucleotides are used to treat, diagnose and prevent immune, developmental and cell signaling disorders and cell proliferative disorders including cancer - Claim 5; Page 181-182; 205pp; English.
XX
XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506. CC CYPR and agonists of CCYPR are used to treat diseases or conditions associated with decreased expression of functional CCYPR, while CCYPR antagonists are used to treat diseases or conditions associated with overexpression of functional CCYPR. Monoclonal or polyclonal antibodies to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect compounds e.g., antibodies, oligonucleotides and proteins (receptors) that specifically bind to CCYPR, and in drug screening methods to identify compounds that modulate the activity of CCYPR. CCYPR nucleotides can be used to generate transgenic animal models of human disease, and can be used in gene therapy in target cells with genetic abnormalities with respect to the expression of CCYPR for the treatment or prevention of a disorder associated with CCYPR. CC Diseases which can be diagnosed, treated and prevented using CCYPR proteins, nucleic acids, agonists or antagonists include immune, developmental and cell signalling disorders, and cell proliferative disorders including cancer. Specific examples of these disorders include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies, diabetes mellitus, disorders of the menstrual cycle and infections caused by bacteria.
XX
XX Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;
SQ
alignment_scores:
Quality: 99.50 Length: 132
Ratio: 1.363 Gaps: 6
Percent Similarity: 55.303 Percent Identity: 29.545
alignment_block:
US-09-327-750D-35 x AAF59611 ..
Align seg 1/1 to: AAF59611 from: 1 to: 898
1 MetAlaSerLysLysGln.....ValIleLeuAspLeu 11
||| |||||||: : : : :
254 ATGGAGTCCAAGAAGACGAGCGCTTAACAATCTCATCGTGGAATAATGT 303
11 uThrValGlulysAspLysLysArgGlyLysAlaSerLysG 28
: : : : : |||: : : : : |||: : : : :
304 CAACCAGAAAATGATGAAAGATGAAAGAGCAGCACTTCATAA. 352
28 InSerGlulGlulProHisLysLeuGlulValGluAsnLysLysPro 44
: : : : : |||: : : : : |||
353 ..GGGAGCCCTTGCCTACCTTTGAATGTTAGTGAATACTGTGCGCT 400
45 GlyGlyAsnValArgArg...LysValArgArgLeuValProAsnPheLe 60
||||| ||||| : : |||||: : : : :
401 AGAGGAACCGTAGCGGTTCCGGCTTAGGCAGGCCATCCTGCAGTATAG 450
60 uTrpAlaIleProAsnArgHisValAsp.....ArgasnG 72
||| ||| : : ||| : : : : :
451 ATGGGACATAATGATGATGCTTGGAGAGCCACAGGCAAGGATGAGAGG 500
72 luGlylGlvGluaSpValGlvArqPheValGlnGlvThrGluVallys 88

```



XX  
PF 21-FEB-2000; 2000EP-0200610.  
XX  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
XX (GEST ) GENSET.  
PA  
XX  
XX  
PI Dumas Milne Edwards J. Duclert A. Giordano J:

XX  
DR WPT: 2000-500381/45.

XX  
PT  
PT  
PT  
XX

XX  
PS  
CLAIM 1; SEQ ID NO.1; /pp + CD-ROM; ENGLISH.  
CC The present sequence is one of a large number of 5' ESTs derived from

mRNAs encoding secreted proteins. An ORF has been identified within the sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolation cDNA sequences derived from the 5' ends of  
CC

CC miRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC miRNAs with intact 5' ends and can therefore be used to obtain full length  
CC

CCDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain

CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
XX

sequence 401 BP; 98 A; 98 C; 148 G; 37 T; 0 other;

alignment\_scores:

Quantity:	52.00	Length:	2
Ratio:	2.359	Caps:	2

correct : foreigner and correct : foreigner

US-09-327-750D-35 x AAC01003 ..

Align seg 1/1 to: AAC01003 from: 1 to: 401

1 MetAlaSerLysLysGlnValIleLeu...AspLeuThrValGluLysAs 16

238 ATGGAGTCCAAAGAGGAACTAGCGGCAACAATCTCAACGGGGAATGC 287

16 pLysLysAspLysArgGly...LysAlaSerLysClnSerGluGluG 32  
.....|.....|.....|.....|.....|.....|

288 CCAACAAGAAACGAAGGAGGGGAGCAGGCCCCACGCAGAATGAAGAAG 337

[illegible]

338 AAATCCCGCCATTTCGGAGGGGGTGAAAGGCCAGAAAGCCCTGGAGGAATATC 387

4 3 2 1 0

[illegible]

\*\*\*\*\*

ID AAI20201 standard; DNA; 276 BP.

AC AAI20201; AC

DT 12-OCT-2001 (first entry)  
yyDE Probe #10134 for gene expression analysis in human cervical cell sample.  
XX



**THIS PAGE BLANK (USPTO)**





```
6024 AGAAACAGGACGACGACGAAGAAGAACACACAAAAAAGAGAGGGA 6073
47 snValArgArgLysValArgLeuValProAsnPheLeuTrpAlaIle 63
6074 AGGAAGAAGGACGAGGAAAAGA..... 6096
64 ProAsnArgHisValasp.....ArgAsnGluGlyGI 74
|||||.....
6097 CCAAGAAAAACACCAACCCGAAAAAGCAGGAGAGAGGACGAGGAGG 6146
74 yGluAspValGlyArgPheValValGlnGlyThrGluValLys..... 88
|||||.....
6147 AAAAGAAGAACCCCAAAAAAACAACGCGACAGACACACAGACCAAA 6196
89 .....ArgLysThrThrGluGlnGlnValArgProTyrArgArgPhe 102
|||||.....
6197 AACAAACACGAAAAACGACGAACAAACACAAAGAGAGAGAAAAAGAAA 6246
103 Arg 103
|||
6247 AGA 6249

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-338-907-179
seq_documentation_block:
; Sequence 179, Application US/09338907
; Patent No. 6265546
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilya, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: PROSTATE CANCER GENE
; FILE REFERENCE: GENSET-18CPICP
; CURRENT APPLICATION NUMBER: US/09/338, 907
; EARLIER FILING DATE: 1999-06-23
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 09/218,207
; EARLIER FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 179
; LENGTH: 56520
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2001..2216
; OTHER INFORMATION: exon1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18196..18265
; OTHER INFORMATION: exon2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 23716..23831
; OTHER INFORMATION: exon3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25570..25659
; OTHER INFORMATION: exon4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 34668..34758
; OTHER INFORMATION: exon5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 40685..40843
; OTHER INFORMATION: exon6
; FEATURE:
```

```
; NAME/KEY: exon
; LOCATION: 48067..48190
; OTHER INFORMATION: exon7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 50179..54519
; OTHER INFORMATION: exon8
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 54493..54498
; OTHER INFORMATION: AATAAA
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1991..2008
; OTHER INFORMATION: upstream amplification primer 5-63
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 2505..2525
; OTHER INFORMATION: downstream amplification primer 5-63 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4091..4111
; OTHER INFORMATION: downstream amplification primer 99-622
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4528..4546
; OTHER INFORMATION: upstream amplification primer 99-622 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5475..5495
; OTHER INFORMATION: downstream amplification primer 99-621
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5927..5947
; OTHER INFORMATION: upstream amplification primer 99-621 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 8127..8144
; OTHER INFORMATION: downstream amplification primer 99-619
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 8560..8578
; OTHER INFORMATION: upstream amplification primer 99-619 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11622..11639
; OTHER INFORMATION: upstream amplification primer 4-76
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12018..12037
; OTHER INFORMATION: downstream amplification primer 4-76 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11930..11947
; OTHER INFORMATION: upstream amplification primer 4-77
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12339..12358
; OTHER INFORMATION: downstream amplification primer 4-77 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12915..12932
; OTHER INFORMATION: upstream amplification primer 4-71
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13317..13334
; OTHER INFORMATION: downstream amplification primer 4-71 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13216..13233
; OTHER INFORMATION: upstream amplification primer 4-72
; FEATURE:
; NAME/KEY: primer_bind
```

```
LOCATION: 13617..13636
OTHER INFORMATION: downstream amplification primer 4-72 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 15994..16011
OTHER INFORMATION: downstream amplification primer 99-610
FEATURE:
NAME/KEY: primer_bind
LOCATION: 16463..16480
OTHER INFORMATION: upstream amplification primer 99-610 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17304..17324
OTHER INFORMATION: downstream amplification primer 99-609
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17814..17832
OTHER INFORMATION: upstream amplification primer 99-609 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18008..18027
OTHER INFORMATION: upstream amplification primer 4-90
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18423..18442
OTHER INFORMATION: downstream amplification primer 4-90 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18699..18716
OTHER INFORMATION: downstream amplification primer 99-607
FEATURE:
NAME/KEY: primer_bind
LOCATION: 19164..19182
OTHER INFORMATION: upstream amplification primer 99-607 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 22589..22609
OTHER INFORMATION: downstream amplification primer 99-602
FEATURE:
NAME/KEY: primer_bind
LOCATION: 23111..23129
OTHER INFORMATION: upstream amplification primer 99-602 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25098..25118
OTHER INFORMATION: downstream amplification primer 99-600
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25657..25674
OTHER INFORMATION: upstream amplification primer 99-600 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 26537..26557
OTHER INFORMATION: downstream amplification primer 99-598
FEATURE:
NAME/KEY: primer_bind
LOCATION: 27022..27040
OTHER INFORMATION: upstream amplification primer 99-598 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32262..32281
OTHER INFORMATION: downstream amplification primer 99-592
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32823..32841
```

```
OTHER INFORMATION: upstream amplification primer 99-592 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34215..34233
OTHER INFORMATION: upstream amplification primer 99-217
FEATURE:

alignment_scores:
Quality: 74.50 Length: 82
Ratio: 1.656 Gaps: 2
Percent Similarity: 54.878 Percent Identity: 30.488

alignment_block:
US-09-327-750D-35 x US-09-338-907-179/rev ..
Align seg 1/1 to reverse of: US-09-338-907-179 from: 1 to: 56520

14 GlutAspLysLysAspLysArgGlyLysAlaSerLysGlnSerG1 30
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39085 GAAAGCAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 39036

30 uGluGluProHisHisLeuGluValGluAsnLysLysProGlyG1A 47
: : : : : : : : : : : : : : : : : : : : : :
39035 AACGAGTGAGGGAAGGAAGAGAGAGGAGGAGGAGAAAGAA 38986

47 snValArgArgLysValArgArgLeuValProAsnPheLeuT1pAla1le 63
: : ||| ||| |||||
38985 GAGAGAGAGCGAAGAAAGAAAGG..... 38963

64 ProAsnArgHisValAspArgAsnGluGlyGlyGlu.....As 76
: : : : : : : : : : : : : : : : : : : : : :
38962 .....AAGAAAGAGGGGAGAGAGACATGAGTTCA 38931

76 pValGlyArgPheValValGlnGlyThrGluValLysArgLysThr 91
: : : : : : : : : : : : : : : : : : : : : :
38930 AATTGGGCATACTGAGGCTGAAGGAACAAACATTTCTAGAAGTAGT 38885

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-996-306-1

seq_documentation_block:
; Sequence 1, Application US/08996306
; Patent No. 5945522
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Chumakov, Ilya
; APPLICANT: Blumefeld, Marta
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,306
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.018A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
```

SEQUENCE CHARACTERISTICS:  
LENGTH: 56516 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: DOUBLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Promoter  
LOCATION: 1629..1870  
IDENTIFICATION METHOD: Proscan  
FEATURE:  
NAME/KEY: Potential ATG  
LOCATION: 1998..2000  
FEATURE:  
NAME/KEY: Exon 1  
LOCATION: 2001..2216  
FEATURE:  
NAME/KEY: ATG  
LOCATION: 2031..2033  
FEATURE:  
NAME/KEY: TYR phos  
LOCATION: 11694..14332  
FEATURE:  
NAME/KEY: SEQ ID42  
LOCATION: 11930..11947  
FEATURE:  
NAME/KEY: SEQ ID24  
LOCATION: 12057..12103  
FEATURE:  
NAME/KEY: SEQ ID51  
LOCATION: compl(12339..12358)  
FEATURE:  
NAME/KEY: SEQ ID64  
LOCATION: 13547..13564  
FEATURE:  
NAME/KEY: SEQ ID58  
LOCATION: 13657..13703  
FEATURE:  
NAME/KEY: SEQ ID67  
LOCATION: compl(13962..13981)  
FEATURE:  
NAME/KEY: Exon 2  
LOCATION: 18196..18265  
FEATURE:  
NAME/KEY: Exon 3  
LOCATION: 23717..23832  
FEATURE:  
NAME/KEY: Exon 4  
LOCATION: 25571..25660  
FEATURE:  
NAME/KEY: SEQ ID43  
LOCATION: 34216..34234  
FEATURE:  
NAME/KEY: SEQ ID25  
LOCATION: 34469..34515  
FEATURE:  
NAME/KEY: SEQ ID52  
LOCATION: compl(34625..34645)  
FEATURE:  
NAME/KEY: Exon 5  
LOCATION: 34669..34759  
FEATURE:  
NAME/KEY: Exon 6  
LOCATION: 40688..40846  
FEATURE:  
NAME/KEY: Exon 7  
LOCATION: 48070..48193  
FEATURE:  
NAME/KEY: Exon 8  
LOCATION: 50182..54523  
FEATURE:

NAME/KEY: SEQ ID65  
LOCATION: 51149..51168  
FEATURE:  
NAME/KEY: SEQ ID59  
LOCATION: 51448..51494  
FEATURE:  
NAME/KEY: SEQ ID68  
LOCATION: compl(51482..51499)  
FEATURE:  
NAME/KEY: SEQ ID44  
LOCATION: 51596..51613  
FEATURE:  
NAME/KEY: SEQ ID26  
LOCATION: 51612..51658  
FEATURE:  
NAME/KEY: SEQ ID53  
LOCATION: compl(51996..52015)  
FEATURE:  
NAME/KEY: polyAd signal  
LOCATION: 54445..54450  
US-08-996-306-1  
  
alignment\_scores:  
Quality: 71.50 Length: 79  
Ratio: 1.663 Gaps: 2  
Percent Similarity: 54.430 Percent Identity: 30.380  
  
alignment\_block:  
US-09-327-750D-35 x US-08-996-306-1/rev ..  
  
Align seg 1/1 to reverse of: US-08-996-306-1 from: 1 to: 56516  
  
17 LysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluPr 33  
||||| ::::| ||||| ||||| ||||| ::::| ::::|  
39077 AAGAAAG 39028  
  
33 OHIsHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArgA 50  
::: ::::| ||||| ||||| ::::| |||||  
39027 GGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38978  
  
50 rGlyValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnArg 66  
||| |||||  
38977 CGAAGAGAAACAGS..... 38964  
  
67 HisValAspArgAsnGluGlyGlu.....AspValGlyAr 79  
:::| ||||| ||||| ::::| ::::|  
38963 .....AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38923  
  
79 gPheValValGlnGlyThrGluValLysArgLysThr 91  
: ::::| ||||| ::::| ::::|  
38922 TACTGAGGCTGAAGAGACAAACATTCTAGAGTAGT 38886  
  
seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-338-907-1  
  
seq\_documentation\_block:  
; Sequence 1, Application US/09338907  
; Patent No. 6265546  
; GENERAL INFORMATION:  
; APPLICANT: Cohen, Daniel  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Ilya, Chumakov  
; APPLICANT: Bougueleret, Lydie  
; TITLE OF INVENTION: PROSTATE CANCER GENE  
; FILE REFERENCE: GENSET.18CP1CP  
; CURRENT APPLICATION NUMBER: US/09/338,907  
; CURRENT FILING DATE: 1999-06-23  
; EARLIER APPLICATION NUMBER: 08/996,306  
; EARLIER FILING DATE: 1997-12-22  
; EARLIER APPLICATION NUMBER: 60/099,658  
; EARLIER FILING DATE: 1998-09-09  
; EARLIER APPLICATION NUMBER: 09/218,207  
; EARLIER FILING DATE: 1998-12-22



```

539 .....ACAGATCGTAA 540
71 nGluGlycylGluAspValGlyArgPheValGlnGlyThrGluValL 88
   :|::|||: |||:||||: |::|||:
550 AAAAGCAGAAGAACGCTTAACCAAAAAGCAGCAGAAAGATAAAGTTA 599
   :|::|||: |||:||||: |::|||:
88 ysArgLysThrThrGluGln...GlnValArgProTyrArgArgPheArg 103
   :|::|||: |||:||||: |::|||:
600 AAAAAAACAGGTGAACAACCACACCGCGCGCTCCAAAAGCAGAA 649
   :|::|||: |||:||||: |::|||:
104 ThrProGluPro 107
   :|::|||: |||:||||: |::|||:
650 AAACCAAGCTCCA 661

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-258-373-21

seq_documentation_block:
; Sequence 21, Application US/09258373
; Patent No. 6150110
; GENERAL INFORMATION:
; APPLICANT: Xiao, Sheng
; APPLICANT: Jonathan A.
; TITLE OF INVENTION: HMG(I/Y)-LAMBA4* FUSION ONCOGENE,
; AND METHODS OF USE
; FILE REFERENCE: B0801/7135/ERP
; CURRENT APPLICATION NUMBER: US/09/258,373
; CURRENT FILING DATE: 1999-02-26
; EARLIER APPLICATION NUMBER: 60/076,401
; EARLIER FILING DATE: 1998-02-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 1875
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-258-373-21

alignment_scores:
Quality: 71.00 Length: 118
Ratio: 1.224 Gaps: 7
Percent Similarity: 49.153 Percent Identity: 29.661

alignment_block:
US-09-327-750D-35 x US-09-258-373-21 ..

Align seg 1/1 to: US-09-258-373-21 from: 1 to: 1875

2 AlasertLysLysGlnValIleLeuAspLeuThrValGluLysAspLys 18
   :|::|||:||||: |||:
244 AGCTCGAAGTCACGACGAGCCCTTGCGCTCCAAAGCAGAAAGGACGGCAC 293
   :|::|||:||||: |||:
18 sApLysArgGly...GlyLysAlaSerLysGlnSerGluGluGluProH 34
   :|::|||:||||: |||:
294 TGAAAGACGGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 343
   :|::|||:||||: |||:
34 iShisLeuGluGluValGluAsnLysLysPro GlyGlyAsnValArgAr 50
   :|::|||:||||: |||:
344 GC.....GAAGTGCACACACTTAAGAGACCTCGGGGGCGGCCAACGAAGGA 387
   :|::|||:||||: |||:
50 gLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnArgH 67
   :|::|||:||||: |||:
388 AGCAAAAACAAAGGTGCTGCCAAGA.....CCGGGAAAC 422
   :|::|||:||||: |||:
67 isValAspArgAsnGluGlyGlyGluAspValGlyArgPheValValGln 83
   :|::|||:||||: |||:
423 CACCACAACCTCCAGGAAGGANAAC.....CAA 448
   :|::|||:||||: |||:
84 GlyThrCluValLys.....ArgLysThrThrGluGlnGlnVa 96
   :|::|||:||||: |||:
449 GGGGCGAGACCCCAAAACATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 498
   :|::|||:||||: |||:
96 lArgProTyrArgPheArgThrPro.....GluProAspAsnHisT 111

```







```

; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"

```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-282-3
;
alignment_scores:
; Quality: 71.00 Length: 71
; Ratio: 1.775 Caps: 4
; Percent Similarity: 56.338 Percent Identity: 33.803
;
alignment_block:
; US-09-327-750d-35 x US-08-402-282-3 ..
;
Align seg 1/1 to: US-08-402-282-3 from: 1 to: 15664
;
49 ArgArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAs 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13436 CGAATCGGTGGCGCAAGGCGCGCGATGCGCGTGG.....CCGAC 13479
;
65 nArg.His.....ValAspArgAsnGluGly 73
|||||:|||||:|||||:|||||:|||||:|||||:
13480 GCGCTCACCCGACCGCGACCTCGAGGTCCGTCGATCGTTCGGGGA 13529
;
74 GlyGluAspValGlyArgPheValGlnGlyThrGluValLysArgly 90
|||||:|||||:|||||:|||||:|||||:|||||:
13530 GGGATGAT.....CGAAGCTACCGCGGCGGACACAGAGAAAAACGGG 13573
;
90 sThrThrGluGlnValArgProTyrArgArgPheArgThrProGluP 107
|||||:|||||:|||||:|||||:|||||:|||||:
13574 GCGACGACGCGAATCGACGCGCCCGGACGACGATCGTGCATC 13623

```

```
107 roAspnHis 110
||
|| .....|||
13624 CTGTTGAGCAT 13634

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-508-004-3

seq_documentation_block:
; Sequence 3, Application US/08508004
; Patent No. 5582969
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Beyer Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,004
; FILING DATE: 27-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,282
; FILING DATE: 10-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 222..425
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
```

```
;
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15429..15664
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
US-08-508-004-3

alignment_scores:
Quality: 71.00 Length: 71
Ratio: 1.775 Gaps: 4
Percent Similarity: 56.338 Percent Identity: 33.803

alignment_block:
US-09-327-750D-35 x US-08-508-004-3 ..
Align seg 1/1 to: US-08-508-004-3 from: 1 to: 15664

49 ArgArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAs 65
||||:||||:||||:||||:||||:||||:||||:||||:||||:
13436 CGAAATCGGTGGCGGCAAGGCGCGCGATGCGGTGG.....CCGAC 13479

65 narg.His.....ValAspArgAsnGluGly 73
:|||||
13480 GGGCTCACCCGACCGGACCTCGAGGTGCGGTGCGGTTCGGGGA 13529

74 GlyGluAspValGlyArgPheValGlnGlyThrGluValLysArgLy 90
||||:|||||
13530 GGGATGAT.....CGAAGCCTACCGCGGACACGAGAAACGCGG 13573

90 sPThrThrGluGlnGlnValArgProTyrArgPheArgThrProGluP 107
||||:||||:||||:||||:||||:||||:||||:||||:
13574 GCGACGACGAGGATCGAGCGCCGACGACGACGATCGTGCATCGATC 13623

107 toAspAsnHis 110
|| :|||
13624 CTGTTTCAGCAT 13634

seq_name: /cgn2_5/ptodata/2/ina/5A_COMB.seq.US-08-402-066-3

seq_documentation_block:
; Sequence 3, Application US/08402066
; Patent No. 5612182
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Bever Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESSEE: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,066
FILING DATE:
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
```

```
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
```

```
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-066-3
;
; alignment_scores:
;   Quality: 71.00      Length: 71
;   Ratio: 1.775       Gaps: 4
;   Percent Similarity: 56.338   Percent Identity: 33.803
;
; alignment_block:
; US-09-327-750d-35 x US-08-402-066-3 ..
;
; Align seg 1/1 to: US-08-402-066-3 from: 1 to: 15664
;
; 49 ArgArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAs 65
; |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; 13436 CGAATCGGTGGCGCAAGGCGCGCGCGATGGCGGTGG.....CCGAC 13479
;
; 65 nArg.His.....ValAspArgAsnGluGly 73
; :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; 13480 GCGCTCACCCCGCAACCGCGACCTCGAGTCCGTCGATCGGTCCGGGA 13529
;
; 74 GlyGluAspValGlyArgPheValGlnGlyThrGluValLysArgLy 90
; |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; 13530 GGGATGAT.....CGAAGCCTACCGCGCGGACACGAGAAACGGCG 13573
;
; 90 sThrThrGluGlnValArgProTyrArgPheArgThrProGluP 107
; : |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 13574 GCGACGACGAGGATCGACGCGCCGACACGACATCGTCACGTCGATC 13623
;
; 107 roAspAsnHis 110
; || ||| |||
; 13624 CTGTCAGCAT 13634
;
; seq_name: /cgn2_6/ptodata/2/lna/5a_comb.seq:US-08-402-068-3
;
; seq_documentation_block:
; ; Sequence 3, Application US/08402068
; ; Patent No. 5633159
; ; GENERAL INFORMATION:
; ; APPLICANT: Pearson, Robert E.
; ; APPLICANT: Dickson, Julie A.
; ; APPLICANT: Hamilton, Paul T.
; ; APPLICANT: Little, Michael C.
; ; APPLICANT: Beyer Jr., Wayne F.
; ; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; ; NUMBER OF SEQUENCES: 6
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ; ADDRESSEE: Company
; ; STREET: 1 Becton Drive
; ; CITY: Franklin Lakes
; ; STATE: NJ
; ; COUNTRY: US
; ; ZIP: 07417
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk
; ; COMPUTER: IBM PC compatible
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: PatentIn Release #1.0, Version #1.25
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/08/402,068
; ; FILING DATE:
; ; CLASSIFICATION: 435
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Fugit, Donna R.
; ; REGISTRATION NUMBER: 32,135
; ; REFERENCE/DOCKET NUMBER: P-3283
; ; INFORMATION FOR SEQ ID NO: 3:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 15664 base pairs
```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 222..425
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
```

```

; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-068-
alignment_scores:
    Quality: 71.00      Length: 71
    Ratio: 1.775      Gaps: 4
    Percent Similarity: 56.338      Percent Identity: 33.803
alignment_block:
    US-09-327-750D-35 x US-08-402-068-3      ..
    Align seg 1/1 to: US-08-402-068-3 from: 1 to: 15664
49 ArgArgLysValArgLeuValProAsnPheLeuTrpAlaIleProAs 65
|||||:|||||
13436 CGAAATCGGTGGCGGCAAGGCGCGGATGCCGTGG.....CGCAC 13479
65 nArg.His.....ValaspArgAsnGluGly 73
|||||
13480 GCGCTCACCCCGAACCCGCGACCTCGAGGTCGCGTTCGGGGGA 13529
```





OM of: US-09-327-750D-35 to: GenEmbl:\* out\_format : pfs  
Date: Mar 11, 2002 3:35 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101154\_20319/app\_query.fasta\_1.1472  
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09327750 -CGNJ=1.8673  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-327-750D-35  
Query length: 117  
Database: GenEmbl:\*  
Database sequences: 1472140  
Database length: -341344837  
Search time (sec): 4557.230000

score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_pr:AK000959	+	355.50	626.28	1.3e-26	1229	AK000959 Homo sapiens cDNA FLJ
gb_pat:AX100231	+	355.50	625.44	1.5e-26	1364	AX100231 Sequence 32 from Pat
gb_pr:HS635C19	+	355.50	593.78	8.5e-25	69648	AL035494 Human DNA sequence
gb_pr:BC003190	+	175.00	306.91	8.1e-09	793	BC003190 Homo sapiens, p75NTR-a
gb_pr:HUMOGC	+	175.00	305.97	9.1e-09	891	M38188 Human unknown protein fr
gb_pr:AF187064	+	175.00	305.97	9.1e-09	891	AF187064 Homo sapiens p75NTR-as
gb_pr:AF187066	+	154.00	270.34	8.8e-07	700	AF187066 Mus musculus p75NTR-as
gb_ro:AR097440	+	154.00	268.74	1.1e-06	854	AF097440 Mus musculus brain exp
gb_scs:G72708	-	150.00	267.28	1.3e-06	421	G72708 MARC 4953-4954:991939031
gb_ro:AF187065	+	149.50	264.70	1.8e-06	519	AF187065 Rattus norvegicus p75N
gb_pr:HSV351F8	+	138.00	208.08	0.0036	45678	AF170719 Human DNA sequence fr
gb_scs:G35294	-	124.50	213.50	0.0013	477	G35294 Human SFS SHGC-37409, se
gb_pr:HS714B7	+	110.50	159.87	1.25	98274	AF097438 Mus musculus brain exp
gb_ro:AR097438	+	107.50	185.74	0.0454	835	AF097438 Mus musculus brain exp
gb_ro:AF051347	+	107.50	185.71	0.0456	838	AF051347 Mus musculus REX-3 mRN
gb_ro:AF097439	+	107.50	177.69	0.1275	2269	AF097437 Mus musculus Bex1 pro
gb_ro:AF097439	+	102.00	176.39	0.1506	785	AF097439 Mus musculus brain exp
gb_pr:AF251053	+	99.50	172.35	0.2528	744	AF251053 Homo sapiens X-linked
gb_pat:AX078272	+	99.50	170.84	0.3070	898	AX078272 Sequence 76 from Paten
gb_pr:HSV870H8	+	99.50	142.25	12.02	31321	G70233 Human DNA sequence fr
gb_pr:AL133348	-	99.50	140.16	15.70	40584	AL133348 Human DNA sequence
gb_hcg:HSU80B1	+	99.50	140.07	15.88	41029	AL022169 Homo sapiens chromos
gb_pr:HS753F9	+	98.50	129.88	58.71	116561	AL023653 Human DNA sequence
gb_hcg:AC012947	+	98.50	125.19	107.09	208618	AF195953 Homo sapiens membra
gb_hcg:AC012947	+	95.00	121.94	162.57	143599	AF012947 Drosophila melanoga
gb_in:AC009183	-	95.00	119.60	219.49	192055	AC009183 Drosophila melanoga
gb_in:AC010124	+	95.00	118.54	221.08	193398	AF010124 Drosophila melanoga
gb_in:AE003685	+	95.00	118.32	258.52	225038	AF003685 Drosophila melanoga
gb_pr:AF183416	+	94.00	163.70	0.7668	642	AF183416 Homo sapiens ovarian g
gb_pr:AF237783	+	94.00	162.02	0.9513	791	AF237783 Homo sapiens brain-exp
gb_pr:AF220189	+	94.00	161.65	0.9972	828	AF220189 Homo sapiens uncharact
gb_pr:HS198P4	+	94.00	131.35	48.61	35714	AF008708 Human DNA sequence
gb_pr:AC002478	+	91.00	120.55	194.16	70113	AC002478 Human BAC clone GSI-
gb_in:AC012332	+	91.00	113.22	497.28	174337	AC012332 Homo sapiens clone
gb_ba:MXAFR2	+	90.50	146.47	6.99	2507	J04157 M.xanthus frzA, frzB an
gb_pr:AP000265	+	90.50	124.52	116.65	38304	AP000265 Homo sapiens genomic
gb_pr:AP000034	-	90.50	116.80	314.19	100000	AP000034 Homo sapiens genomic
gb_pr:AP000101	+	90.50	116.80	314.19	100000	AP000101 Homo sapiens genomic
gb_pr:AP000177	+	90.50	116.80	314.19	100000	AP000177 Homo sapiens genomic
gb_hcg:AC026930	+	90.50	114.66	413.39	130443	AC026930 Homo sapiens chromid

gb\_pr:AC005295 + 90.50 114.62 415.44 131070 ! AC005295 Homo sapiens xp2  
gb\_pr:AP001713 - 90.50 106.95 1.1e+03 340000 ! AP001713 Homo sapiens gen  
gb\_ba:AE008371 - 90.00 133.00 39.36 11967 ! AE008371 Agrobacterium tum  
gb\_pr:U72787 + 90.00 129.48 61.75 18510 ! U72787 Homo sapiens cosmid  
gb\_pr:AL353806 + 90.00 115.06 392.66 111051 ! AL353806 Human DNA sequen

seq\_name: gb\_pr:AK000959

seq\_documentation\_block: 1229 bp mRNA PRI 22-FEB-2000  
LOCUS AK000959  
DEFINITION Homo sapiens cDNA FLJ10097 fis, clone HEMB1002458, weakly similar  
to OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74.

ACCESSION AK000959

VERSION AK000959.1 GI:7021945

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to

mrna, clone\_lib:HEMBA1 clone:HEMBA1002458.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (sites)  
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,  
Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,  
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,  
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A.

NEDO human cDNA sequencing project

Unpublished (2000)

2 (bases 1 to 1229)

Isogai,T. and Otsuki,T.

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao

Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,

Kisarazu, Chiba 292-0812, Japan [E-mail:genomics@hri.co.jp,

Tel:81-438-52-3951, Fax:81-438-52-3952]

NEDO human cDNA sequencing project supported by Ministry of

International Trade and Industry of Japan; cDNA full insert

sequencing: Research Association for Biotechnology; cDNA library

construction, 5'- & 3'-end one pass sequencing and clone selection;

Helix Research Institute (supported by Japan Key Technology Center

etc.) and Department of Virology, Institute of Medical Science,

University of Tokyo.

FEATURES

Location/Qualifiers

1..1229

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HEMBA1002458"

/dev\_stage="embryo, 10 weeks"

/tissue\_type="whole embryo, mainly head"

/note="cloning vector: pME18SFL3"

191..553

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BAA91443.1"

/db\_xref="GI:7021946"

/translation="MESSELAANLNGENAGQNEGEQAPQTONEERSHLGGGEGO

KPGGNIRGRVRLVPNRPWAI PNPHIEHNEARDDVRFVGMELIKRKTRDQQRHY

MRQTPEDPDNDYFCLIP"

BASE COUNT 351 a 243 c 296 g 339 t

ORIGIN

..

alignment\_scores:

Quality: 355.50 Length: 120

Ratio: 3.665 Gaps: 3

Percent Similarity: 80.833 Percent Identity: 59.167

alignment\_block:

US-09-327-750D-35 x AK000959

Align seg 1/1 to: AK000959 from: 1 to: 1229

```

1 MetAlaSerLysLysGlnValIleLeu...AspLeuThrValGluLysAs 16
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 ATGAGTTCACAAAGAGGAAGTCTAGCGGCAACAACTCTCAACGGGAAAATGC 240

16 pLysLysAspLysArgGlyGly...LysAlaSerLysGlnSerGluGluG 32
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 CCAACAAGAAACGAGGAGGGGAGCAGGCCCCACCCAGCAATGAAGAAG 290

32 luProHisHisLeuGluGluValGluValGluAsnLysLysProGlyGlyAsnVal 48
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 AATCCCGCCATTTGGAGGGGGTGAAGGCCAGAGCCCTGGAGGAATATC 340

49 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 AGCGGGGGCGAGTTAGCGGACCTGTCCCTAAATTTTCGATGGGCCATACC 390

64 oAsnArgHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheV 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
391 TAATAGGCATATTGAGCACAAATCAAGCGAGAGATGATGTAGAAGGTTTG 440

81 alValGlnGlyThrGluValLysArgLysThrThrGluGlnGlnValArg 97
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
441 TAGGCCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATCAGG 490

98 ProTyrArgArgPheArgThrProGluProAspAsnHisTyrAspPheCy 114
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
491 CACTATATGCGCTTCCAAACTCTGAACTGACACCTGACACCATTTATGACTTTG 540

114 sLeuIlePro 117
|||||:|||||:
541 CCTCATACCT 550

```

seq\_name: gb\_pat:AX100231

```

seq_documentation_block:
LOCUS AX100231 1364 bp DNA PAT 02-APR-2001
DEFINITION Sequence 32 from Patent WO0119860.
ACCESSION AX100231
VERSION AX100231.1 GI:13539115
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Tang,Y.T., Hillman,J.L., Yue,H., Reddy,R., Lal,P., Shah,P.,
Azimzai,Y., Baughin,M.R., Lu,D.A., Bandman,O., Shih,L.L. and
Patterson,C.
TITLE Proteins associated with cell differentiation
JOURNAL Patent: WO 0119860-A 32 22-MAR-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
location/Qualifiers
1..1364
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 1990956CBI"
BASE COUNT 411 a 269 c 322 g 362 t
ORIGIN

```

```

alignment_scores:
Quality: 355.50 Length: 120
Ratio: 3.665 Gaps: 3
Percent Similarity: 80.833 Percent Identity: 59.167

```

alignment\_block:

US-09-327-750D-35 x AX100231 ..

Align seg 1/1 to: AX100231 from: 1 to: 1364

```

1 MetAlaSerLysLysGlnValIleLeu...AspLeuThrValGluLysAs 16
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 ATGAGTTCACAAAGAGGAAGTCTAGCGGCAACAACTCTCAACGGGAAAATGC 299

```

```

16 pLysLysAspLysArgGlyGly...LysAlaSerLysGlnSerGluGluG 32
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 CCAACAAGAAACGAGGAGGGGAGCAGGCCCCACCCAGCAATGAAGAAG 349

32 luProHisHisLeuGluGluValGluValGluAsnLysLysProGlyGlyAsnVal 48
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
350 AATCCCGCCATTTGGAGGGGGTGAAGGCCAGAGCCCTGGAGGAATATC 399

49 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 AGCGGGGGCGAGTTAGCGGACCTGTCCCTAAATTTTCGATGGGCCATACC 449

64 oAsnArgHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheV 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 TAATAGGCATATTGAGCACAAATCAAGCGAGAGATGATGTAGAAGGTTTG 499

81 alValGlnGlyThrGluValLysArgLysThrThrGluGlnGlnValArg 97
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
500 TAGGCCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATCAGG 549

98 ProTyrArgArgPheArgThrProGluProAspAsnHisTyrAspPheCy 114
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
550 CACTATATGCGCTTCCAAACTCTGAACTGACACCTGACACCATTTATGACTTTG 599

114 sLeuIlePro 117
|||||:|||||:
600 CCTCATACCT 609

```

seq\_name: gb\_pr:HS635G19

seq\_documentation\_block:

```

LOCUS HS635G19 69648 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 635G19 on chromosome Xq22.1-22.3
Contains a LAMRI (Laminin Receptor 1 (67kD) (RPSA, 40S Ribosomal
Protein SA, P40)) pseudogene and part of a novel protein. Contains
ESTs and GSSs, complete sequence.
ACCESSION AL035494
VERSION AL035494.8 GI:4775622
KEYWORDS HTG; 40S Ribosomal Protein P40; 40S Ribosomal Protein SA; CpG
Island; Laminin Receptor 1; LAMRI; RPSA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 69648)
JOURNAL Direct Submission
Submitted (02-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

```

COMMENT

On May 11, 1999 this sequence version replaced gi:4678749.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome X, constructed by the Sanger Centre Chromosome X  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/ChrX  
635G19 is from the library RPC14 constructed at the Roswell Park  
Cancer Institute by the group of Pieter de Jong. For further

details see <http://bacpac.med.buffalo.edu/VECTOR:pcYPAC2>  
IMPORTANT: This sequence is not the entire insert of clone 635G19.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone U65A4 (281014) is at 69549 in this  
sequence. The true right end of clone U101D3 (285997) is at 100 in  
this sequence.

## FEATURES

Source	Location/Qualifiers
	1..69648
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="q22.1-22.3"
	/clone="RP4-635G19"
	/clone_lib="RPC1-4"
repeat_region	439..483
	/note="TARI repeat: matches 707..749 of consensus"
repeat_region	501..661
	/note="AluB repeat: matches 137..311 of consensus"
repeat_region	662..897
	/note="118 copies 2 mer aa 70% conserved"
repeat_region	915..1331
	/note="MSTA repeat: matches 1..419 of consensus"
repeat_region	1334..1517
	/note="92 copies 2 mer ga 77% conserved"
repeat_region	1927..2020
	/note="MIR repeat: matches 109..217 of consensus"
repeat_region	2066..2262
	/note="LIME2 repeat: matches 5825..6022 of consensus"
repeat_region	2263..2574
	/note="AluSp repeat: matches 1..313 of consensus"
repeat_region	2575..2819
	/note="LIME2 repeat: matches 5583..5825 of consensus"
repeat_region	2821..3039
	/note="Alu repeat: matches 1..311 of consensus"
repeat_region	3104..3384
	/note="LIME2 repeat: matches 5304..5591 of consensus"
repeat_region	3391..4007
	/note="LIMB1 repeat: matches 5558..6162 of consensus"
repeat_region	4012..4137
	/note="Trigger2a repeat: matches 11..123 of consensus"
repeat_region	4141..4252
	/note="L1 repeat: matches 2920..3031 of consensus"
repeat_region	4265..4350
	/note="HY4 repeat: matches 1..87 of consensus"
repeat_region	4350..4678
	/note="MER74A repeat: matches 28..369 of consensus"
repeat_region	4914..5095
	/note="MERSA repeat: matches 8..189 of consensus"
repeat_region	5224..5555
	/note="L2 repeat: matches 2423..2746 of consensus"
repeat_region	6040..6086
	/note="L2 repeat: matches 1676..1720 of consensus"
repeat_region	7413..7478
	/note="Alu repeat: matches 239..304 of consensus"
repeat_region	7628..7928
	/note="AluB8 repeat: matches 1..316 of consensus"
repeat_region	7952..7987
	/note="18 copies 2 mer aa 81% conserved"
repeat_region	7990..8305
	/note="AluJo repeat: matches 1..309 of consensus"
repeat_region	8436..8538
	/note="MIR repeat: matches 99..212 of consensus"
repeat_region	10836..11442
	/note="LIMD repeat: matches 1357..1972 of consensus"
repeat_region	11443..11571
	/note="FLAM A repeat: matches 2..129 of consensus"
repeat_region	11572..11946
	/note="LIMD repeat: matches 977..1357 of consensus"
repeat_region	11947..12245
	/note="AluY repeat: matches 1..301 of consensus"
repeat_region	12246..13098
	/note="match: GSS B63622"
repeat_region	13099..13223
	/note="L1MD repeat: matches 275..977 of consensus"
repeat_region	13523..13820
	/note="L1MD repeat: matches -17..113 of consensus"
repeat_region	13823..13932
	/note="MLTID repeat: matches 177..505 of consensus"
repeat_region	13936..13977
	/note="55 copies 2 mer ct 77% conserved"
repeat_region	13978..14240
	/note="MLTID repeat: matches 201..241 of consensus"
repeat_region	14245..14534
	/note="AluSg1 repeat: matches 1..270 of consensus"
repeat_region	14535..14721
	/note="AluSc repeat: matches 1..290 of consensus"
repeat_region	14724..15604
	/note="MLTID repeat: matches 2..201 of consensus"
repeat_region	15605..15909
	/note="LIMC/D repeat: matches 4646..5565 of consensus"
repeat_region	16069..16363
	/note="AluSc repeat: matches 24..312 of consensus"
repeat_region	16482..16641
	/note="AluX repeat: matches 1..291 of consensus"
repeat_region	16671..16953
	/note="AluB repeat: matches 136..295 of consensus"
repeat_region	17164..17280
	/note="AluX repeat: matches 1..284 of consensus"
repeat_region	17326..17377
	/note="L1P repeat: matches 4499..4619 of consensus"
repeat_region	17704..17863
	/note="26 copies 2 mer at 79% conserved"
repeat_region	17864..18175
	/note="AluSc repeat: matches 118..295 of consensus"
repeat_region	18176..18298
	/note="AluYb8 repeat: matches 1..318 of consensus"
repeat_region	18848..18873
	/note="AluSc repeat: matches 1..118 of consensus"
repeat_region	18876..19171
	/note="13 copies 2 mer tg 100% conserved"
repeat_region	19183..19232
	/note="AluSg repeat: matches 1..295 of consensus"
repeat_region	19246..19533
	/note="25 copies 2 mer ac 100% conserved"
repeat_region	19539..19837
	/note="AluY repeat: matches 5..292 of consensus"
repeat_region	20016..20194
	/note="MER74B repeat: matches 61..368 of consensus"
repeat_region	20394..20692
	/note="MER74A repeat: matches 295..487 of consensus"
repeat_region	20701..20809
	/note="AluSg repeat: matches 1..302 of consensus"
repeat_region	20877..21061
	/note="LTR1 repeat: matches 682..785 of consensus"
repeat_region	21186..21636
	/note="LTR29 repeat: matches 1..195 of consensus"
repeat_region	21609..21737
	/note="LTR29 repeat: matches 3..498 of consensus"
repeat_region	21771..22076
	/note="MER34 repeat: matches 409..543 of consensus"
repeat_region	22878..22974
	/note="AluY repeat: matches 2..308 of consensus"
repeat_region	23452..23582
	/note="MIR repeat: matches 92..191 of consensus"
repeat_region	23583..23624
	/note="FLAM_C repeat: matches 1..131 of consensus"
repeat_region	23735..24041
	/note="21 copies 2 mer ta 76% conserved"
repeat_region	24045..24072
	/note="AluSg1 repeat: matches 1..308 of consensus"
repeat_region	25274..25582
	/note="14 copies 2 mer tt 89% conserved"
repeat_region	25583..25747
	/note="AluB repeat: matches 1..289 of consensus"
misc_feature	25748..25747

```
repeat_region 25959..26012
/Note="HERVPH21 repeat: matches 23..79 of consensus"
repeat_region 26144..26456
/Note="AluJ repeat: matches 1..311 of consensus"
repeat_region 27281..27432
/Note="AluJb repeat: matches 137..289 of consensus"
repeat_region 27436..27507
/Note="L1M4 repeat: matches 5360..5434 of consensus"
repeat_region 27536..27600
/Note="L1M4 repeat: matches 2220..2283 of consensus"
repeat_region 27601..27975
/Note="MLT1A1 repeat: matches 3..365 of consensus"
repeat_region 27976..28273

alignment_scores:
  Quality: 355.50      Length: 120
  Ratio: 3.665        Gaps: 3
  Percent Similarity: 80.833  Percent Identity: 59.167

alignment_block:
US-09-327-750D-35 x HS635G19 ..

Align seg 1/1 to: HS635G19 from: 1 to: 69648

1 MetAlaSerLysLysGlnValIleLeu...AspLeuThrValGluLysAs 16
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64293 ATGGAGTCCAAAGAGGAAGTACCGCGCAACAATCTCAACGGGAAATGC 64342

16 pLysLysAspLysArgGlyGly...LysAlaSerLysGlnSerGluGluG 32
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
64343 CCAACAAGAAACCAAGAGGGAGGAGCGAGCGCCACGACAGAAATGAAGAAG 64392

32 LuProHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnVal 48
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
64393 AATCCCGCCATTTGGGAGGGGGTGAAGCCAGAGCCTGGAGGAATATC 64442

49 ArgArg...LysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 64
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
64443 AGCGGGGGCGAGTTAGCGCACTGTCTCCTAATTTTCGATGGGCCATACC 64492

64 oAsnArgHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheV 81
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
64493 TAATAGGCATATTGAGCAACAATGAAGCAGAGATGATGTAAGAAGTTG 64542

81 alValGlnGlyThrGluValLysArgLysThrThrGluGlnGlnValArg 97
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
64543 TAGGCCAGATGATGTAAGTAAGAAAGACTAGGGAACACGACATGAGG 64592

98 ProTyrArgArgPheArgThrProGluProAspAsnHisTyrAspPheCy 114
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
64593 CACTATATGGCTTCCAAACTCTGAACCTGACACCACTTATGACTTTTG 64642

114 sLeuIlePro 117
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
64643 CCTCATACCT 64652

seq_name: gb_pr:BC003190

seq_documentation_block:
LOCUS BC003190 793 bp mRNA PRI 12-JUL-2001
DEFINITION Homo sapiens, p75NTR-associated cell death executor; ovarian
granulosa cell protein (13kd), clone MGC:802 IMAGE:3357965, mRNA,
complete cds.
ACCESSION BC003190
VERSION BC003190.1 GI:13112030
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 793)
AUTHORS Strausberg,R.
TITLE Direct Submission
```

```
JOURNAL Submitted (13-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Seedi, Jacqueline
Scheln, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Scott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 6 Row: k Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7657043.

FEATURES
  source
    1..793
      /organism="Homo sapiens"
      /db_xref="LocusID:27018"
      /db_xref="taxon:9606"
      /clone="MGC:802 IMAGE:3357965"
      /tissue_type="Eye, retinoblastoma"
      /clone_lib="NIH_MGC_16"
      /lab_host="DH10B-R"
      /note="Vector: pOTB7"
      196..531
        /codon_start=1
        /product="p75NTR-associated cell death executor; ovarian
granulosa cell protein (13kd)"
        /protein_id="AAH03190.1"
        /db_xref="GI:13112031"
        /translation="MANIHOENEMEQMNCEDRPLGGEGEHOPAGNRRGQARELA
PNFRWAIPNRQINDMGDGDDEMFMEHREIRKRLRELQLRNLRLMGLSNNHHD
HHDEFCLMP"
      BASE COUNT 225 a 170 c 205 g 193 t
      ORIGIN

alignment_scores:
  Quality: 175.00      Length: 121
  Ratio: 2.273        Gaps: 4
  Percent Similarity: 63.636  Percent Identity: 36.364

alignment_block:
US-09-327-750D-35 x BC003190 ..

Align seg 1/1 to: BC003190 from: 1 to: 793

4 LysLysGlnValIleLeu...AspLeuThrValGluLysAspLysLysAs 19
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
181 AAAAAAATCTCATTCATGCGCAATATTTCCACGAGGAACCAAGAGATGGA 230

19 pLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluProHisHisL 36
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
231 GCAG.....CCTATGCAGAAATGGAGAGAGAGACGCGCCTT 265

36 euGluGluValGluAsnLysLysProGlyGlyAsnValArgArgLysVal 52
||||| :|||||:|||||:|||||:|||||:|||||:|||||:
266 TGGGAGAGGTGAAGGCCACCGCTTCAGGAATCGACGGGACAGGCT 315

53 ArgArgLeuValProAsnPheLeuTrpAlaIleProAsnArgHisValas 69
```

```

|||||.....|||||
316 CGCGGACTTCCCTTAATTTTCGATGGCCATACCCAATAGGAGATCAA 365
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
69 pArgAsnGluGly.....GlyGluAspValGlyArgPheValValGlnG 84
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
366 TGATGGATGGTGGAGATGGAGATGATATGGAATATTCATGAGAGAGA 415
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
84 lyThrGluVallyAsArgLysThrThrGluGlnGlnValArgProTyArg 100
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
416 TGAGAGAAATCAGAGAAATTTAGGAGAGCTGAGTTGAGCAATTCGTCG 465
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
101 ArgPheArgThrProGluProAspAsnHisTyArg.....Ph 113
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
466 CGTATCCTTATGGGGAGCTCTTAATCACCATGACCATCATGATCAATT 515
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
113 eCysLeuIlePro 117
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
516 TTGCCTTATGCCT 528

```

seq\_name: gb\_pr:HUMOGC

```

seq_documentation_block:
LOCUS HUMOGC 891 bp mRNA PRI 07-MAR-1995
DEFINITION Human unknown protein from clone PHGR74 mRNA, complete cds.
ACCESSION M38188 X56942
VERSION M38188.1 GI:189378
KEYWORDS
SOURCE Human ovarian granulosa cell line, cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Rapp,G., Freudenstein,J., Klaudiny,J., Mucha,J., Wempe,F.,
Zimmer,M. and Scheit,K.H.
TITLE Characterization of three abundant mRNAs from human ovarian
granulosa cells
JOURNAL DNA Cell Biol. 9 (7), 479-485 (1990)
MEDLINE 91025550
COMMENT Draft entry and computer-readable sequence for [DNA 9, 479-485
(1990)] kindly submitted
by K.H.Scheit, 27-AUG-1990.
FEATURES
source
Location/Qualifiers
1..891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PHGR74"
/cell_type="granulosa"
/tissue_type="ovary"
<1..891
/note="protein of unknown function"
312..647
/note="protein of unknown function"
/codon_start=1
/protein_id="AAA63232.1"
/db_xref="GI:189379"
/translation="MANIHQNEEMEQPMQGEEDRPLGGEGHOPAGNRRGQARRLA
PNFRWAIPIRQINDCMGGDDMEIFMEEMREIRKRLQLRNLRLMGLSNHHD
HHDEFCLMP"
361..534
/note="protein of unknown function"
/codon_start=1
/protein_id="AAA63233.1"
/db_xref="GI:189380"
/translation="MERKALWEEVKATSLQIDGDLADLPLIFDGPYPITGRSMGW
VEMEWIKYSRR"
BASE COUNT 251 a 182 c 224 g 234 t
ORIGIN

```

alignment\_scores:  
Quality: 175.00 Length: 121  
Ratio: 2.273 Gaps: 4  
Percent Similarity: 63.636 Percent Identity: 36.364

alignment\_block:

US-09-327-750D-35 x HUMOGC

Align seg 1/1 to: HUMOGC from: 1 to: 891

```

4 LysLysGlnValIleLeu...AspLeuThrValGluLysAspLysLysAs 19
|||||.....|||||
297 AAAAAAATCTCATTCATGCGCAATATTCACGAGGAAAAACGAGAGATGGA 346
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
19 pLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluProHisL 36
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
347 GCAG.....CCTATGCAGAAATGAGAGGAAAGACCGCCCTT 381
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
36 euGluGluValGluAsnLysLysProGlyGlyAsnValArgArgLysVal 52
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
382 TGGGAGGAGGTGAAGCCACCGCTGCAGAAATCGACGGGACAGGCT 431
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
53 ArgArgLeuValProAsnPheLeuTrpAlaIleProAsnArgHisValas 69
|||||.....|||||
432 CGCGGACTTCCCTTAATTTTCGATGGCCATACCCAATAGGAGATCAA 481
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
69 pArgAsnGluGly.....GlyGluAspValGlyArgPheValValGlnG 84
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
482 TGATGGATGGTGGAGATGGAGATGATATGGAATATTCATGAGAGAGA 531
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
84 lyThrGluVallyAsArgLysThrThrGluGlnGlnValArgProTyArg 100
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
532 TGAGAGAAATCAGAGAAATCTTAGGAGGTGCAGTTGAGGAATTCGTCG 581
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
101 ArgPheArgThrProGluProAsnHisTyArg.....Ph 113
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
582 CGTATCCTTATGGGGAGCTCTCTAATCACCATGACCATCATGATGAATT 631
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
113 eCysLeuIlePro 117
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
632 TTGCCTTATGCCT 644

```

seq\_name: gb\_pr:AF187064

```

seq_documentation_block:
LOCUS AF187064 891 bp mRNA PRI 11-JUN-2000
DEFINITION Homo sapiens p75NTR-associated cell death executor (NADE) mRNA,
complete cds.
ACCESSION AF187064
VERSION AF187064.1 GI:8452893
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,
Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,I.A. and Sato,T.A.
TITLE NADE, a p75NTR-associated cell death executor, is involved in
signal transduction mediated by the common neurotrophin receptor
p75NTR
JOURNAL J. Biol. Chem. 275 (23), 17566-17570 (2000)
MEDLINE 20298829
REFERENCE
AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
Hanaoka,T., Li,Y., Irie,S. and Sato,T.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
University, 630 West 168th St., P&S II-451, New York, NY 10032, USA
FEATURES
Location/Qualifiers
1..891
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..891
/gene="NADE"
312..647
/gene="NADE"

```



```
115 euLlePro 117
|||||
541 TTATGCCT 548

seq_name: gb_ro:AF097440

seq_documentation_block:
LOCUS AF097440 854 bp mRNA ROD 13-APR-1999
DEFINITION Mus musculus brain expressed x-linked protein 3 (Bex3) mRNA,
complete cds.
ACCESSION AF097440
VERSION AF097440.1 GI:4580593
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 854)
Brown,A.L. and Kay,G.F.
Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
Hum. Mol. Genet. 8 (4), 611-619 (1999)
99172070
Erratum: [[published erratum appears in Hum Mol Genet 1999
May;8(5):943]]
2 (bases 1 to 854)
Brown,A.L. and Kay,G.F.
Direct Submission
TITLE
Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
JOURNAL
Location/Qualifiers
FEATURES
source
1..854
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/db_xref="dbEST:AA272375"
/chromosome="X"
/map="near Plp"
/tissue_type="pooled organs"
1..854
/gene="Bex3"
172..546
/gene="Bex3"
/codon_start=1
/product="brain expressed x-linked protein 3"
/protein_id="AAD24431.1"
/db_xref="GI:4580594"
/translation="MANVHOENEEMEQPLONGEDRPVGCCEGHQPAANNNNNNHNN
HNHRRQARLAPNFRWAIPIRNMNDGLGGDDMEMFMEERIRKRLQLRNC
LRILMGLSNHHDDHDFCLMP"
BASE COUNT 237 a 212 c 228 g 177 t
ORIGIN

alignment_scores:
Quality: 154.00 Length: 136
Ratio: 1.925 Gaps: 4
Percent Similarity: 58.824 Percent Identity: 30.147

alignment_block:
US-09-327-750d-35 x AF097440 ..
Align seg 1/1 to: AF097440 from: 1 to: 854
1 MetAlaSerLysLysGlnValIleLeuAspLeuThrValGluLysAspLys 17
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145 ATAGCAACCAAGGAAACAAATCTCATCGGCCCAATGTCACCAAGGAAA 194
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 sLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluProH 34
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
195 CGAAGAGATG.....GAGCAGCCCTCGCAGATGGACAGAGACC 235
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 iShisLeuGluGluValGluAsnLysLysProGlyGlyAsn..... 47
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

seq_name: gb_sts:G72708
seq_documentation_block:
LOCUS G72708 421 bp DNA STS 08-AUG-2001
DEFINITION MARC 4953-4954:991939031.1 SCF - porcine spleen Sus scrofa STS
genomic, sequence tagged site.
ACCESSION G72708
VERSION G72708.1 GI:15146738
KEYWORDS STS.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 421)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L. and
Keeler,J.W.
Single nucleotide polymorphism (SNP) discovery in expressed porcine
genes
Unpublished (2001)
JOURNAL
COMMENT
Contact: Freking BA
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4278
Fax: 402 762 4173
Email: freking@email.marc.usda.gov
Primer A: GCAATGGGTGAAACTCTACT
Primer B: CCAGCAGCAATAGACG
STS size: 500
PCR profile:
Hotstart: 95 degrees for 15 minutes
Denature: 95 degrees for 30 seconds
Anneal: 56 degrees
Extension: 68 degrees for 2 minutes
Cycles: 32 to 45
Protocol:
Template: 50-200 ng genomic DNA
Primer: each 20 pmoles
dNTPs: each 88 uM
Taq Polymerase: 0.25 units (Qiagen HotStar)
Buffer:
Commercially supplied Qiagen HotStar buffer
The STS is derived from PCR amplicons generated from genomic DNA,
sequenced from each end using the amplification primers. The
```





LOCUS HSV351F8 45678 bp DNA PRI 23-NOV-1999  
 DEFINITION Human DNA sequence from cosmid V351F8, between markers DXS366 and DXS87 on chromosome X contains ESTs.  
 ACCESSION Z70719  
 VERSION Z70719.1 GI:1261915  
 KEYWORDS X.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 45678)  
 Whiteley, M.  
 Direct Submission  
 Submitted (09-APR-1995) Sanger Centre, Hinxton, Cambridgeshire, CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk  
 JOURNAL  
 COMMENT IMPORTANT: This sequence is the entire insert of clone V351F8. The true left end of clone V351F8 is at 1 in this sequence. The true right end of clone V351F8 is at 45678.  
 V351F8 is from the human chromosome X-specific cosmid library.  
 FEATURES  
 Location/Qualifiers  
 1..45678  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="X"  
 /chromosome="X"  
 /clone="GHC-351F8"  
 /clone\_lib="SCCV"  
 1..70  
 /partial  
 /note="Alu repeat: matches 80. .1 of consensus"  
 642..930  
 /note="Alu repeat: matches 1. .308 of consensus"  
 1406..5179  
 /note="L1 element fragment"  
 5243..5276  
 /note="17 copies of 2 mer 82 & conserved"  
 5279..5567  
 /partial  
 /note="Alu repeat: matches 304. .1 of consensus"  
 9975..10369  
 /note="MSTA element fragment"  
 10399..12017  
 /note="MSTAR element fragment"  
 11318..11978  
 /note="THE1BR element fragment"  
 11715..11846  
 /note="MLTIR element fragment"  
 12074..12170  
 /note="MSTA element fragment"  
 12227..12486  
 /note="MSTA element fragment"  
 12437..12502  
 /note="MSTC element fragment"  
 12756..13059  
 /partial  
 /note="Alu repeat: matches 308. .1 of consensus"  
 14478..14650  
 /note="THE1B element fragment"  
 14491..14536  
 /note="MSTA element fragment"  
 14715..14831  
 /note="THE1B element fragment"  
 15099..15533  
 /note="L1 element fragment"  
 15534..15626  
 /note="31 copies of 3 mer 85 & conserved"  
 15566..15619  
 /note="3 copies of 18 mer 98 & conserved"  
 16352..16436  
 /note="L1 element fragment"  
 16728..16763  
 /note="9 copies of 4 mer 94 & conserved"  
 16764..17054

/partial  
 /note="Alu repeat: matches 308. .1 of consensus"  
 19420..19449  
 /note="15 copies of 2 mer 87 & conserved"  
 complement(19537..20013)  
 /note="match: 3' EST N51315 clone 283089"  
 20071..20114  
 /note="22 copies of 2 mer 98 & conserved"  
 20073..20112  
 /note="10 copies of 4 mer 100 & conserved"  
 20073..20126  
 /note="3 copies of 18 mer 87 & conserved"  
 23088..23300  
 /note="L1 element fragment"  
 23311..23490  
 /note="MLT1A element fragment"  
 23318..23489  
 /note="MLT1B element fragment"  
 23362..23491  
 /note="MLT1C element fragment"  
 23507..24407  
 /note="L1 element fragment"  
 24376..24462  
 /note="MSTC element fragment"  
 24379..24491  
 /note="MSTA element fragment"  
 24688..24738  
 /note="MSTA element fragment"  
 24810..25016  
 /note="L1 element fragment"  
 25254..25421  
 /note="L1 element fragment"  
 25439..25696  
 /note="MER25 element fragment"  
 26284..26343  
 /note="MLT1D element fragment"  
 26428..26491  
 /note="MLT1D element fragment"  
 26465..26518  
 /note="MLT1E element fragment"  
 27446..27667  
 /note="L1 element fragment"  
 28954..29247  
 /partial  
 /note="Alu repeat: matches 308. .1 of consensus"  
 30160..30325  
 /note="2 copies of 83 mer 98 & conserved"  
 32588..32875  
 /note="match: 5' EST H68599 clone 239077"  
 complement(32825..33230)  
 /note="match: 3' EST H68239 clone 289077"  
 35551..35802  
 /partial  
 /note="Alu repeat: matches 1. .260 of consensus"  
 35804..35839  
 /note="18 copies of 2 mer 83 & conserved"  
 38625..38705  
 /note="MLT1R element fragment"  
 38747..38839  
 /note="MLT1R element fragment"  
 38949..39032  
 /note="MLT1E element fragment"  
 38965..39076  
 /note="MLT1D element fragment"  
 BASE COUNT 14260 a 9135 c 9327 g 12956 t  
 ORIGIN

alignment\_scores:  
 Quality: 138.00 Length: 127  
 Ratio: 2.000 Gaps: 4  
 Percent Similarity: 54.331 Percent Identity: 31.496

## alignment\_block:

US-09-327-750D-35 x HSV351F8 ..

Align seg 1/1 to: HSV351F8 from: 1 to: 45678

11 LeuThrValGluLysAspLysLysAspLysArgGlyGlyLysAlaSerLy 27  
 19408 TTATCTTCTTAGAAGAAAAAACAACAGGAAAAAACAACAT 19457  
 27 sGlnSerGluGluProHisHisLeuGluGlu.....ValGluA 41  
 19458 GGAATAATGTCCCAAGGAAAAACAAGTTGTGGAGAGGCCCACTGCAGA 19507  
 41 sn.....LysLysProGlyGlyAsn 47  
 19508 ATGAGCCCCCGCTTTTAGGAGGTGGTAATACAGGAGCCTGGAGGAAT 19557  
 48 ValArgArgLysValArgArgLeuValProAsnPheLeuTrpAlaIlePr 64  
 19558 GTTAAAGGGGTTTGGGCTCCACCTGCCCGGGTTTGGAGAGGATGTGCC 19607  
 64 oAsnArgHisValAspArg.....AsnGluGlyGlyGluAspV 77  
 19608 CAATAGCCTTGTGCATACATTCATATGATAGATGGAGATGGAGATGATA 19657  
 77 alGlyArgPheValValGlnGlyThrGluValLysArgLysThrThrGlu 93  
 19658 TGGACGGTTCATGGAGGAGATGAGAGCTAAGGAGGAAATAGGGAA 19707  
 94 GlnGlnValArgProTyrArgArgPheArgThrProGluPro..... 107  
 19708 CTTCAAGTTGAGGTAGCTGCGCATCTTATAGGGAGACCTCCTCACCA 19757  
 108 AspAsnHisTyrAspPheCysLeuIlePro 117  
 19758 TGATCATCATGATGATGAGTTTGCCTTATGCCT 19788

seq\_name: gb\_sts:G35294

seq\_documentation\_block: 477 bp DNA 02-OCT-1997  
 LOCUS human STS SHGC-37409, sequence tagged site.  
 DEFINITION G35294  
 ACCESSION G35294  
 VERSION G35294.1 GI:2459462  
 KEYWORDS STS; STS sequence; primer; sequence tagged site.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 477)  
 MYERS R.M.  
 JOURNAL Unpublished (1997)  
 COMMENT

Contact: Richard M. Myers  
 Stanford Human Genome Center (SHGC)  
 Stanford University School of Medicine  
 Department of Genetics, M-344, Stanford, CA 94305, USA  
 Tel: 4157259687  
 Fax: 4157259689  
 Email: myers@shgc.stanford.edu

Primer A: AACATCTTTCATGAAAGTTGATG

Primer B: CTTTGGCATCTTCTGCAG

STS size: 106

PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 15 seconds

Annealing: 60 degrees C for 30 seconds

Polymerization: 72 degrees C for 23 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9600

Protocol:

Template: 25 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 Amplifraq Gold Polymerase: 0.07 units/uL  
 Total Vol: 5 uL

Buffer: MgCl2: 2.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from N51315  
 -- Washington University/Merck EST sequence.

## FEATURES

source  
 1..477  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 STS  
 44..149  
 primer\_bind  
 44..67  
 primer\_bind  
 complement(130..149)  
 BASE COUNT 153 a 115 c 79 g 130 t  
 ORIGIN

## alignment\_scores:

Quality: 120.50 Length: 83  
 Ratio: 2.410 Gaps: 2  
 Percent Similarity: 60.241 Percent Identity: 36.145

## alignment\_block:

US-09-327-750D-35 x G35294/rev ..

Align seg 1/1 to reverse of: G35294 from: 1 to: 477

42 LysLysProGlyGlyAsnValArgArgLysValArgArgLeuValProAs 58  
 474 CAGGAGCCTGGAGGAATGTTAAAGGGTTTGGCTCCACCTGCCCGGG 425  
 58 nPheLeuTrpAlaIleProAsnArgHisValAspArg.....A 71  
 424 TTTTGGAGAGCATGTGCCCAATAGGCTTGTGCAATACATTTGATATGATAG 375  
 71 snGluGlyGlyGluAspValGlyArgPheValValGlnGlyThrGluVal 87  
 374 ATGAGATGAGATGATATGGAACGGTTTCATGGAGGAGATGAGAGAGCTA 325  
 88 LysArgLysThrThrGluGlnGlnValArgProTyrArgArgPheArgTh 104  
 324 AGGAGGAATTTAGGGAACCTTCAGTTGAGGTACAGCTCGGCATTTCTTAT 275  
 104 rProGluPro.....AspAsnHisTyrAspPheCysLeuIlePro 117  
 274 AGGGACCTTCCTCACCATCATCATCATGATGATGATGATGATGATGATG 226

seq\_name: gb\_pr:HS714B7

## seq\_documentation\_block:

LOCUS HS714B7 98274 bp DNA PRI 12-DEC-1999  
 DEFINITION Human DNA sequence from clone CTA-714B7 on chromosome 22q12.2-13.2  
 Contains pseudogene similar to part of COX7B (cytochrome c oxidase subunit VIIb), a novel pseudogene, ESTs, STS and GSSs, complete sequence.

ACCESSION 299755  
 VERSION 299755.1 GI:3036782  
 KEYWORDS HTG; COX7B; cytochrome c oxidase.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 98274)  
 CONNOR, R.  
 Direct Submission



/note="MIR repeat: matches 102. .143 of consensus"  
25378. .25555  
/note="L2 repeat: matches 2330. .2519 of consensus"  
25673. .25910  
/note="MIR repeat: matches 4. .255 of consensus"  
26722. .26798  
/note="L2 repeat: matches 2572. .2649 of consensus"  
27040. .27107  
/note="L2 repeat: matches 2489. .2500 of consensus"  
27108. .27374  
/note="AluY repeat: matches 41. .304 of consensus"  
27375. .27471  
/note="L2 repeat: matches 2372. .2489 of consensus"  
27651. .27945  
/note="AluX repeat: matches 1. .299 of consensus"  
28235. .28876  
/note="L2 repeat: matches 1113. .1817 of consensus"  
28975. .29097  
/note="MER96 repeat: matches 42. .173 of consensus"  
29173. .29196  
/note="12 copies 2 mer aa 100 conserved"  
29832. .30260  
/note="match: GSS: Em:AQ075670"  
30392. .30597  
/note="MIR repeat: matches 35. .252 of consensus"  
30845. .31010  
/note="MER91B repeat: matches 2. .162 of consensus"  
31056. .31094  
/note="MIR repeat: matches 218. .256 of consensus"  
31728. .31791  
/note="16 copies 4 mer tgtg 76 conserved"  
31858. .31952  
/note="MIR repeat: matches 48. .153 of consensus"  
complement(33437. .33589)  
/note="match: StS: Em:H55731"  
33543. .33633  
/note="MIR repeat: matches 84. .165 of consensus"  
33785. .33966  
/note="MER91A repeat: matches 1. .185 of consensus"

alignment\_scores:  
Quality: 114.50 Length: 66  
Ratio: 2.602 Gaps: 3  
Percent Similarity: 66.667 Percent Identity: 39.394

alignment\_block:

US-09-327-750D-35 x HS714B7 ..

Align seg 1/1 to: HS714B7 from: 1 to: 98274

55 LeuValProAsnPheLeuTtPAlaIleProAsnArgHisValAsp..... 69  
||||:|||||  
62887 TTGGCCCTTAATTTTCAATGGGCGCATACCAATAGGCGAGTCAATGATGG 62936  
||||:|||||  
70 ArgAsnGluGlyGlyGluAspValGlyArgPheValValcInGlyThrG 86  
||||:|||||  
62937 GATCAATAGAGTGGAGATGATATGTAATGTAATGTTTCATGAGGAGATGAG 62986  
||||:|||||  
86 luValLysArgLys.....ThrThrGluGlnGlnValArgPro 98  
||||:|||||  
62987 GAATCAGGAGAGATAAGGAGTACAAATTCAGGAATGTCTGTATPCTT 63036  
||||:|||||  
99 TyrArgArgPheArgThrProGluProAspAsnHisTyrAspPheCys 114  
||||:|||||  
63037 TATGGGAAGCTGTGTAATCCCAT...GACCATCATGATGAATTTTGC 63081  
||||:|||||

seq\_name: gb\_ro:AF097438

seq\_documentation\_block:  
LOCUS AF097438 835 bp mRNA ROD 11-APR-1999  
DEFINITION Mus musculus brain expressed x-linked protein 1 (Bex1) mRNA,  
complete cds.  
ACCESSION AF097438

AF097438.1 GI:4580589  
house mouse.  
Mus musculus  
Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Mus.  
1 (bases 1 to 835)  
Brown, A. L. and Kay, G. F.  
Bex1, a gene with increased expression in parthenogenetic embryos,  
is a member of a novel gene family on the mouse X chromosome  
Hum. Mol. Genet. 8 (4), 611-619 (1999)  
99172070  
Erratum: [[published erratum appears in Hum Mol Genet 1999  
May;8(5):943]]  
2 (bases 1 to 835)  
Brown, A. L. and Kay, G. F.  
Direct Submission  
Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of  
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia  
Location/Qualifiers  
1..835  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/map="near plp"  
/tissue\_type="blastocyst"  
1..835  
/gene="Bex1"  
205..591  
/gene="Bex1"  
/codon\_start=1  
/product="brain expressed X-linked protein 1"  
/protein\_id="AAD24429.1"  
/db\_xref="GI:4580590"  
/translation="MESKDKGVKNLNMENDHOKBEKEKPODTIRREPAVALISEAG  
KNCAPRGRRFRVROPJAHYRWDLQMRVGPQGRMREENVQFGDVRQLMEKLRR  
QLSHSLRAVSTDPHPHHHDEPCLMP"  
BASE COUNT 238 a 184 c 246 g 167 t  
ORIGIN

alignment\_scores:  
Quality: 107.50 Length: 129  
Ratio: 1.473 Gaps: 6  
Percent Similarity: 56.589 Percent Identity: 30.233

alignment\_block:

US-09-327-750D-35 x AF097438 ..

Align seg 1/1 to: AF097438 from: 1 to: 835

1 MetAlaSerLysLysGlnValIleLeuAspLeuThrValGluLysAsp.. 16  
||||:|||||  
205 ATGAGTCCAAAGATCAAGCGCTGAAATCTCAACATGGAGAATGACCA 254  
||||:|||||  
17 ....LysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluG 32  
||||:|||||  
255 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 304  
||||:|||||  
32 luProHisHisLeu.....GluGluValGluAsnLysLysPro... 44  
||||:|||||  
305 AGCAGGTGTGGCCCTCGATCTCCGAGGCTGGCAAAACTGTGCGCTAGA 354  
||||:|||||  
45 GlyClyAsnValArgArgLysValArgArgLeuValProAsnPheLeuTr 61  
||||:|||||  
355 GGAGGTGCGAGGCGGTTCGCGGTTCGCGCAGCCCATCGCTCACTATAGTG 404  
||||:|||||  
61 pAlaIleProAsnArgHisValAspArgAsnGluGly.....GlyG 75  
||||:|||||  
405 GGACCTGATCGCAGG...GTGGGGAGCCCCAGGGAAGGATGAGAGAGG 451  
||||:|||||  
75 luAspValGlyArgPheValValcInGlyThrGluValLysArgLysThr 91  
||||:|||||

**THIS PAGE BLANK (USPTO)**



```

33 roHishisLeuGluValGluAsnLysProGlyGlyAsnValArg 49
|||||
246 CCCACCATCGGAAGAGTTGAAACAAGAGCTGGGGAAATGTCCGA 295
|||||
50 ArgLysValArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 66
|||||
296 AGGAAAGTCAGGCGACTTGCTGCTTCTCTGGGCCATACCAATAG 345
|||||
66 ghisValAspArgAsnGluGlyGlyGluAspValGlyArgPheValValG 83
|||||
346 GCATGTTGATCGCAATGAGGGGGAGAGCATGTTGGGAGATTGTCGTGC 395
|||||
83 InGlyThrGluValLysArgLysThr 91
|||||
396 AGGGAACAGAGTCAAGAGAAAGACT 421
|||||

```

seq\_name: gb\_est1:AW251884

```

seq_documentation_block:
LOCUS      AW251884      540 bp      mRNA      EST      17-DEC-1999
DEFINITION UI-R-BJ0-adn-b-06-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone
            UI-R-BJ0-adn-b-06-0-UI 3', mRNA sequence.
ACCESSION  AW251884
VERSION    AW251884.1 GI:5595475
KEYWORDS   EST.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 540)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   9704477
COMMENT   Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized atrium at 16.5 dpc library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA-res.
            Location/Qualifiers
            1..540
               /organism="Rattus norvegicus"
               /strain="Sprague-Dawley"
               /db_xref="taxon:10116"
               /clone="UI-R-BJ0-adn-b-06-0-UI"
               /dev_stage="adult"
               /lab_host="DH10B (Life Technologies)"
               /note="Vector: pMT3D-Pac (Pharmacia) with a modified
               polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0
               library is a subtracted library derived from the UI-R-AAL,
               UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
               UI-R-AG1 libraries. These libraries represent tissues from
               rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
               at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
               canal at 15 dpc, and ventricle at 13 dpc. The tag is a
               string of 5-6 nucleotides present between the Not I site
               and the oligo-dT track. The library was constructed as
               described by Bonaldo, Lennon and Soares, Genome Research

```

#### FEATURES

source

```

6: 791-806, 1996.
TAG_LIB=UI-R-BJ0
TAG_TISSUE=atrium at 16.5 dpc
TAG_SEQ=GATTC"
BASE COUNT      146 a      120 c      94 g      180 t
ORIGIN
alignment_scores:
  Quality: 459.00      Length: 97
  Ratio: 5.044      Gaps: 0
  Percent Similarity: 93.814      Percent Identity: 86.598
alignment_block:
US-09-327-750D-35 x AW251884/rev ..
Align seg 1/1 to reverse of: AW251884 from: 1 to: 540
21 ArgGlyGlyLysAlaSerLysGlnSerGluGluGluProHisHisLeuCl 37
:::|||||
538 AAAGGTGGGAGGCGCTCCAAACAAGTGAAGAGAATCCACCATCTGGA 489
|||||
37 uGluValGluAsnLysLysProGlyGlyAsnValArgLysValArgA 54
|||||
488 ACAGGTTGAAAACAAGACCGCTGGGGCAATGTGAGGAGGAAGTCAGGC 439
|||||
54 rGluValProAsnPheLeuTrpAlaIleProAsnArgHisValAspArg 70
|||||
438 GACTTGTGCCTAACTTCTATGGGCCATACCTAATAGGCATGTTGATCAC 389
|||||
71 AsnGluGlyGlyGluAspValGlyArgPheValValGlnGlyThrGluVa 87
:::|||||
388 ACTGAAGGGGGAGAGAGGCTTGGGAGATTCTGAGGCAGGTGATGGAAGC 339
|||||
87 lLysArgLysThrThrGluGlnGlnValArgProTyrArgArgPheArgT 104
:::|||||
338 CAAGAGAAAGTCTAAGGAGCAACAGATGAGGCTTACACGCGTTCCCGAA 289
|||||
104 hrProGluProAspAsnHisTyrAspPheCysLeuIlePro 117
|||||
288 CCCCTGAACCTGACAATCATATTAGGACTTTTGCTCATACCT 248
|||||

```

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 465)

REFERENCE  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, I., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Willie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE  
JOURNAL  
COMMENT  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through ILNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:229324

Seq primer: ETPPrimer  
High quality sequence stop: 328.

#### FEATURES

Location/Qualifiers  
1..465  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:367892"  
/clone\_lib="Soares mouse embryo NBMEl3.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/note="Vector: pMT3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGCGCGCGGAAATTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 147 a 90 c 156 g 72 t  
ORIGIN

alignment\_scores:  
Quality: 469.50 Length: 99  
Ratio: 4.840 Gaps: 1  
Percent Similarity: 97.980 Percent Identity: 96.970

#### alignment\_block:

us-09-327-750d-35 x W54487 ..

Align seg 1/1 to: W54487 from: 1 to: 465

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGlulLysAs 16  
|||||  
168 ATGGCATCCAAATTTAAACAAGTCATACCTGGATCTCCTGAGGAGAAGA 217  
|||||  
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluP 33  
|||||  
218 CAAAAAGACAAAAGGTGGGAGGCCCTCCAAACAAGTGAAGAAGAAC 267  
|||||  
33 roHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49  
|||||  
268 CCCACCATCTGGAAGAGGTGAAACAAGAAGCCTGGGGGAATGTCCGA 317  
|||||  
50 ArgLysValIleArgLeuValProAsnPheLeuTirPAlaIleProAsnAr 66  
|||||

318 AGAAGTC AGCGACTTGTGCTTAACATTTCTCTGGGCCCATACCAATAG 366  
66 gHisValAspArgAsnGlyGlyGlyGluAspValGlyArgPheValValC 83  
|||||  
367 GCATGTTGATCGCAATGAGGGGGAGAGGATGTTGGGAGATTGTAGTGC 416  
|||||  
83 InGlyThrGluValLysArgLysThrThrGluGlnGlnValValArgPro 98  
|||||  
417 AGGACACAGAAGTCAGAGAAAGACTACGAGCAGCAGGTGAGGCCT 463  
seq\_name: gb\_est1:AL118340

seq\_documentation\_block:  
LOCUS AL118340 422 bp mRNA EST 23-SEP-1999  
DEFINITION v9112b26 Beddington mouse dissected endoderm Mus musculus CDNA  
clone 528\_13N20 5', mRNA sequence.

ACCESSION AL118340  
VERSION AL118340.1 GI:5920179  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 422)  
AUTHORS Harrison, S.M., Dunwoodie, S.L., Arkell, R.M., Lehrach, H. and Beddington, R.S.  
TITLE Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissue constituents of the gastrulating mouse embryo  
JOURNAL Development 121 (8), 2479-2489 (1995)  
MEDLINE 95401865  
COMMENT Contact: Ruiz, P., Lehrach, H. and Avner, P.  
EEC Mouse Transcript Mapping Consortium  
Genoscope - CNS  
2, rue Gaston Cremieux, CP 5706, 91057 Evry CEDEX, France  
Email: pavner@pasteur.fr  
clone available from Ressourcenzentrum, Heubnerweg 6, D-14059 Berlin, Germany. Web site http://www.rzpd.de  
Seq primer: CCGTCGCGAATTCGCGGT.  
Location/Qualifiers  
1..422  
/organism="Mus musculus"  
/strain="C57B16 x DBA"  
/db\_xref="taxon:10090"  
/clone="528\_13N20"  
/clone\_lib="Beddington mouse dissected endoderm"  
/tissue\_type="dissected endoderm"  
/dev\_stage="7.5 dpc"  
/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. ~ High quality sequence only submitted. ~ Average insert size: 1.2 kb (range: 0.2 - 2.kb)"

BASE COUNT 128 a 86 c 143 g 65 t  
ORIGIN

#### alignment\_scores:

Quality: 463.50 Length: 92  
Ratio: 5.093 Gaps: 1  
Percent Similarity: 98.913 Percent Identity: 98.913

#### alignment\_block:

us-09-327-750d-35 x AL118340 ..

Align seg 1/1 to: AL118340 from: 1 to: 422

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGlulLysAs 16  
|||||  
146 ATGGCATCCAAATTTAAACAAGTCATACCTGGATCTCCTGAGGAGAAGA 195  
|||||  
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluP 33  
|||||  
196 CAAAAAGACAAAAGGTGGGAGGCCCTCCAAACAAGTGAAGAAGAAC 245  
|||||



```

100 ArgArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeu1116
302 ACCGTTTTCGAACCCCTGAACCTGACATCATACGACTTTTGCCTCAT 253
116 ePro 117
1111
252 ACCT 249

```

seq\_name: gb\_est1:AA117429

```

seq_documentation_block:
LOCUS AA117429 312 bp mRNA EST 15-NOV-1996
DEFINITION mn2le12.r1 Beddington mouse embryonic region Mus musculus cDNA
clone IMAGE:538606 5' similar to SW:HG74_HUMAN Q00994 OVARIAN
GRANULOSA CELL 13.0 KD PROTEIN HGR74. ;, mRNA sequence.
ACCESSION AA117429
VERSION AA117429.1 GI:1672442
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le.M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

```

```

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

```

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:325542

Seq primer: -40m13 ET

High quality sequence stop: 287.

Location/Qualifiers

source

```

1. .312
/organism="Mus musculus"
/strain="C57BL6 x DBA"
/db_xref="taxon:10090"
/clone="IMAGE:538606"
/clone_lib="Beddington mouse embryonic region"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="7.5dpc"
/lab_host="DH12S"
/Note="Organ: whole embryo; Vector: pCMV-Sport; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Gastrulating embryos were collected at 7.5dpc
from C57BL6 x DBA matings, excluding embryos that had
developed head folds and all extraembryonic tissues.
Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
Referenced in Development 121, 2479-2489 (1995)."
BASE COUNT 110 a 56 c 92 g 54 t
ORIGIN

```

alignment\_scores:

```

Quality: 503.50 Length: 100
Ratio: 5.086 Gaps: 1
Percent Similarity: 99.000 Percent Identity: 98.000

```

alignment\_block:

US-09-327-750D-35 x AA117429 ..

Align seg 1/1 to: AA117429 from: 1 to: 312

```

1 MetAlSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
|||||
12 ATGCATCCAAATTTAAACAAGTCATCTGATCTCACTCTGGAGAAGA 61
|||||
16 pLysLysAspLysArgGlyLysAlaSerLysGlnSerGluGluGlu 33
|||||
62 CAAAAAGACAAAAAGGTGGAGGCTCCAAACAAGTGAAGAAGAAC 111
|||||
33 roHisHisLeuGluValGluAsnLysLysProGlyGlyAsnValArg 49
|||||
112 CCCACCATCTGGAAGAGTTGAAACAAAGAGCTGGGGAAATGTCGGA 161
|||||
50 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 66
|||||
162 AGAAAGTCAGCGACTTGTGCCTAACTTCTCTGGGCCATACCTAATAG 211
|||||
66 gHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheValG 83
|||||
212 GCATGTTGATCGCAATGAAGGGGAGAGGATGTTGGGAGATTTGTAGTC 261
|||||
83 InGlyThrGluValLysArgLysThrThrGluGlnGlnValArgProTyr 99
|||||
262 AGGAACAGAAAGTCAGAGAAAGACTACGAGCAGCAGGTCAGGCCCTAC 311
|||||
seq_name: gb_est2:W35893

```

seq\_documentation\_block:

```

LOCUS W35893 472 bp mRNA EST 14-MAY-1996
DEFINITION mc53909.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:352288 5' similar to SW:HG74_HUMAN Q00994 OVARIAN
GRANULOSA CELL 13.0 KD PROTEIN HGR74. [1] ;, mRNA sequence.
ACCESSION W35893
VERSION W35893.1 GI:1317733
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

REFERENCE

AUTHORS

```

1 (bases 1 to 472)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le.M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

```

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:224088

Seq primer: EPrimer

High quality sequence stop: 441.

Location/Qualifiers

source

1. .472

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:352288"

/clone\_lib="Soares mouse embryo NBME13.5 14.5"

/sex="unknown"

/tissue\_type="embryo"

/dev\_stage="13.5-14.5dpc total fetus"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'

```

|||||
428 CCCACCATCTCGAAGAGGTTGAAACAAGAGCCTGGGGCAATGTGTCAGG 379
|||||
50 ArgLysValArgArgLeuValProAsnPhenLeuTrpAlaIleProAsnAr 66
|||||
378 AGGAAAGTCAGGCGACTGTGCGCTAACTTTCTATGGGCCATACCTAATAG 329
|||||
66 gHisValAspArgAsnGluGlyGluAspValGlyArgPheValValG 83
|||||
328 GCATGTTGATCAGTCAGTGAAGGGGAGAGGTTGGAGAGATTCGTAGGGC 279
|||||
278 AGGTGATGGAAGCCAGAGAAAGTCTAAGGAGCAACAGATGAGCGCTTAC 229
|||||
83 lnglyThrGluValLysArgLysThrGluGlnGlnValArgProTyr 99
|||||
100 ArgArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeuII 116
|||||
228 ACCGTTTCGAACCCCTGAACCTGAACATCAATAGGACTTTTGCCCTCAT 179
|||||
116 ePro 117
|||||
178 ACCT 175

```

seq\_name: gb\_est2:BI289546

```

seq_documentation_block: 643 bp mRNA EST 19-JUL-2001
LOCUS BI289546
DEFINITION UI-R-DK0-cff-c-12-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone
UI-R-DK0-cff-c-12-0-UI 3', mRNA sequence.
ACCESSION BI289546
VERSION BI289546.1 GI:14947228
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 643)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

```

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized rat kidney pool library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) the following repetitive elements were found in this cDNA sequence: 1-51,  
>AT\_rich#Low\_complexity  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES  
source

```

Location/Qualifiers
1..643
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DK0-cff-c-12-0-UI"
/clone_lib="UI-R-DK0"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DK0

```

library is a subtracted library derived from a mixture of five individually tagged normalized rat libraries: brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%), aorta-nRAP (20%), and placenta-nRPP (20%). Each original library was constructed from a mixture of equal amounts of RNA from seven different developmental time-points: embryonic day 17, embryonic day 19, embryonic day 21, adult day 1, adult day 12, adult day 75, and adult day 200. (Exception: the aorta pool does not contain embryonic day 17 RNA and the placenta pool contains only the three embryonic stages). Each library was normalized individually according to the procedure described by Bonaldo, Lennon & Soares (Genome Research Genome 6: 791-806, 1996). For construction of the DK0 subtracted library, plasmid DNA from each of the five individually tagged normalized libraries was mixed in the proportions specified above and electroporated into competent bacteria for production of single-stranded circular DNA representing the pool of libraries. Single-stranded circular DNA representing these five normalized libraries was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a set of about 1,000 arrayed clones from each of the five non-normalized libraries of brain (CT0s), heart (CS0s), kidney (CU0s), aorta (CW0s), and placenta (CX0s). The resulting pool of approximately 5,000 clones represented about 33.3% of the final driver population. A set of about 2,000 arrayed clones from each of the five normalized libraries of brain (CT0), heart (CS0), kidney (CU0), aorta (CW0), and placenta (CX0). The resulting pool of about 10,000 clones represented about 66.6% of the final driver population.

BASE COUNT 160 a 149 c 113 g 221 t  
ORIGIN  
TAG-TISSUE-rat kidney pool  
TAG\_SEQ=CAAGACTGTC"

alignment\_scores:  
Quality: 511.50 Length: 118  
Ratio: 4.608 Gaps: 1  
Percent Similarity: 94.068 Percent Identity: 86.441

alignment\_block:  
US-09-327-750d-35 x BI289546/rev ..

Align seg 1/1 to reverse of: BI289546 from: 1 to: 643

```

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16
|||||
601 ATGGCGCTCCAAAGTCAACAAGTCATACCTGGATCTCCTGTCGGAAGA 552
|||||
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluLup 33
|||||
551 CAAGAAAAACAAAAAGGTGGGAG.GCCTCCAAACAAGTGAAGAGAAT 503
|||||
33 rHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49
|||||
502 CCCACCATCTGGAAGAGGTTGAAAACAAGAGCCTGGGGCAATGTGTCAGG 453
|||||
50 ArgLysValArgArgLeuValProAsnPhenLeuTrpAlaIleProAsnAr 66
|||||
452 AGGAAAGTCAGGCGACTGTGCTAACTTTCTATGGGCCATACCTAATAG 403
|||||
66 gHisValAspArgAsnGluGlyGluAspValGlyArgPheValValG 83
|||||
402 GCATGTTGATCAGTCAGTGAAGGGGAGAGGTTGGAGAGATTCGTAGGGC 353
|||||
83 lnglyThrGluValLysArgLysThrGluGlnGlnValArgProTyr 99
|||||
352 AGGTGATGGAAGCCAGAGAAAGTCTAAGGAGCAACAGATGAGCGCTTAC 303

```



(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M.Fatima

BASE COUNT		ORIGIN	
158 a	109 c	170 g	81 t
		1 others	

alignment_scores:		
Quality:	568.50	Length: 113
Ratio:	5.168	Gaps: 1
Percent Similarity:	97.345	Percent Identity: 96.460

alignment\_block:

US-09-327-750D-35 x BE723075

Align seq 1/1 to: BF723075 from: 1 to: 519

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAsp 16  
181 ATGGCATCCAAATTTAAACAAGTCATCTGGATCTCACTGTGAGAAGA 230  
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluP 33  
231 CAAAAGAACA AAAAGTGGGAGGCGCTCCAAACAAGTGAAGAAGAAC 280

33 roliHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49  
281 CCCACCATCTGGAACAGGTTGAAACAAAGAGCCCTGGGGAAATGTCGGA 330  
50 ArgLysValArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 66  
331 AGGAAAGTCAGCGCACTTGTGCCTAACTTCTCTGGGCCCACTCAATAG 380

66 gHisValAspArgAsnGluGlyGluAspValGlyArgPheValValG 83  
381 GCATGTTCATGCCAATGAAGGGGAGAGATGTTGGGAGATTTGTASTGC 430  
83 InGlyThrGluValLysArgLysThrThrGluGluGlnValArgProTyr 99  
431 AGGGAAACAGANTCAAGAGAAGAAGACTACGGAGCAGCAGGTGAGCGCTTAC 480

100 ArgArgPheArgThrProGluProAspAsnHisTyrAsp 112  
|||||  
481 AGCGGTTCGCAACGCCGGACCTGCACATCATTCGAC 519  
|||||

seq\_name: qb\_est1:BE333774

seq_documentation_block:	
LOCUS	BE333774 463 bp mRNA EST 14-JUL-2000
DEFINITION	us27h06.y1 Soares_NNEBA_branchial_arch Mus musculus cDNA clone IMAGE:3168347 5' similar to SW:HC74_HUMAN Q00994 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74... mRNA sequence.

ACCESSION	BE333774
VERSION	BE333774.1
KEYWORDS	EST.

SOURCE	ORGANISM
house mouse.	Mus musculus

REFERENCE  
AUTHORS  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

**JOURNAL**  
**COMMENT**  
Unpublished (1997)  
Contact: Robert Strausberg.

CONTACT: ROBERT C. STAUDSBERG, Ph.D.  
Email: cgapops-remail.nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
, Ph.D.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:

[image.llnl.gov/image/html/resources.shtml](http://image.llnl.gov/image/html/resources.shtml)

MGI:1063807

Seq primer: -40RP from GIBCO.

FEATURES	source
----------	--------

/organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3168347"  
 /clone\_lib="Soares\_NWEBA\_branchial\_arch"  
 /tissue\_type="branchial arches"  
 /dev\_stage="embryo, 10.5 dpc"  
 /lab\_host="DH10B (phage resistant)"  
 /note="vector: pT73D-Pac (Pharmacia) w/  
 polylinker; Site\_1: NotI; Site\_2: EcoRI;  
 was primed with a Not I - oligo(dT) prim  
 TGTTCACAACTGCAAGTGGAGCGCGCATGATTTT  
 3']; double-stranded cDNA was ligated to  
 (pharmacia), digested with Not I and cl  
 I and Eco RI sites of the modified pT77  
 constructed and normalized by Bento Soar  
 Ronaldo."

BASE COUNT	155 a	91 c	147 g	70 t
ORIGIN	BONAIRO.			

alignment_scores:		
Quality:	544.50	Length: 113
Ratio:	4.950	Gaps: 1
Percent Similarity:	97.345	Percent Identity: 95.575

alignment block:

US-09-327-750D-35 X BE333774

Align seq 1/1 to: BE3333774 from: 1 to: 463

1 MetAlaSerLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16  
126 ATGGCATCCAAATTAAACAAGTCATCTGGATCTCACTGTGGAGAAGA 175  
16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGlnGluP 33  
176 CAAANAAGACAAAAGGTGGGAGGCCCTCCAAACAAGTCGAGAAGAAC 225  
33 roHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49  
226 CCACCACTTCGACAGAGTTGAACAAGAACGCTCGGGGAATGTCGGA 275

50 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 66  
276 AGGAAACCTCCGACACTTGTGCCTAACCTTCTCTGGGCCAATAC 325  
66 gHisValAspArgAsnGluGlyGlyGluAspValGlyArgHisVal 83  
326 GCATCTTTCGCCAATGAAGGGGAGGAGACATGTTGGGAGATTTC 374

83 InClgYThrGluValLysArgIysThrThrGluGlnGlnValArgYrotyr 99  
375 AGGGAAACAGAGACTCAAGAGAAGACTACGGAGCAGCAGGTGAGGCCTTAC 424  
100 ArgArgPheArgThrProGluProAspAsnHisTyrAsp 112  
425 AAGCGTTTTCCGAACCCCGGAACCTGCACAATCATTTACGAC 463

seq\_name: qb\_est2:W17712

sea documentation block:

seq_documentation_block:	EST	10-SEP-1996
LOCUS	W17712	317 bp
DEFINITION	mb77e04.r1 Soares mouse p3NWF19.5 Mus musculus cDNA clone	
IMAGE	IMAGE:335454 5' similar to SW:HC74_HUMAN Q00994 OVARIAN GRANU	
CELL	CELL 13.0 KD PROTEIN HGR74. [1] ; mRNA sequence.	
ACCESSION	W17712	
VERSION	W17712.1	GI:1292113



This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI-511536

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 417.

#### FEATURES

source

1..458

/organism="Mus musculus"

/strain="BALB/c"

/db\_xref="taxon:10090"

/clone="IMAGE:872056"

/clone\_lib="Barstead MPLRB1"

/sex="mixed"

/tissue\_type="kidney"

/dev\_stage="6 weeks"

/lab\_host="DH10B"

/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTACGAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTT 3'); double-stranded cDNA was ligated to Eco RI adaptors [CATGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bob Barstead."

BASE COUNT 147 a 88 c 115 g 108 t

ORIGIN

#### alignment\_scores:

Quality: 605.00 Length: 116

Ratio: 5.261 Gaps: 0

Percent Similarity: 99.138 Percent Identity: 97.414

#### alignment\_block:

US-09-327-750D-35 x AA473525 ..

Align seg 1/1 to: AA473525 from: 1 to: 458

2 AlaserLysLysGlnValIleLeuAspLeuThrValGluLysAspLysLy 18

|||||

13 GCCAAATTTAAACAAGTCATCTGATCTCACTGTGGAGAAAGACAAAA 62

18 sAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluProHis 35

|||||

63 ACACAAAAAGCTGGGAAGGCTCCAAACAAGTGAGAGAAGACCCACC 112

35 isLeuGluGluValGluAsnLysLysProGlyGlyAsnValArgArgLys 51

|||||

113 ATCTGGAAGAGCTTGAACAAAGAAGCTGGGGAATGTCCGAAGGAAA 162

52 ValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnArgHisVa 68

|||||

163 GTCAGGCGACTTGTGCCCTAACTTCTCTGGGCCATACCAATAGGCATGT 212

68 lAspArgAsnGluGlyGlyGluAspValGlyArgPheValGlnGlyT 85

|||||

213 TGATCGGCANTGAAGGGGAGAGGATTTGGGAGATTTGTAGTCAGGGAA 262

85 hrGluValLysArgLysThrThrGluGlnGlnValArgProTyrArgArg 101

|||||

263 CAGAAGTCAACAGAAAGACTACCGAGCAGCAGGTGAGGCCTTACAGGCGT 312

102 PheArgThrProGluProAsnHisTyrAspPheCysLeuIlePro 117

|||||

313 TTCGGAACCCCGAACCTGACAATCATTTACGACTTTTTCCTCATACCT 360

seq\_name: gb\_est2:BG070341

seq\_documentation\_block:

LOCUS BG070341 747 bp mRNA EST

DEFINITION H3086C08-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone

ACCSSION BG070341 H3086C08 3', mRNA sequence.

#### alignment\_scores:

Quality: 611.50 Length: 118

Ratio: 5.226 Gaps: 1

Percent Similarity: 99.153 Percent Identity: 98.305

#### alignment\_block:

US-09-327-750D-35 x BG083261 ..

Align seg 1/1 to: BG083261 from: 1 to: 792

1 MetaLaserLys...LysGlnValIleLeuAspLeuThrValGluLysAs 16

|||||

190 ATGGCATCCAAATTTAAACAAGTCATCTGATCTCACTGTGGAGAAAGA 239

16 pLysLysAspLysArgGlyGlyLysAlaSerLysGlnSerGluGluGluP 33

|||||

240 CAAAAAGACAAAAAAGGTGGGAAGGCTCCAAACAAGTGAAGAAGAAC 289

33 roHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 49

|||||

290 CCCACCATCTCGAAGAGGTGGAACAAGAGCCTGGGGGAATGTCCGA 339

50 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 66

|||||

340 AGGAAAGTCAGCGACTTGTGCCTAACTTCTCTGGGCCATACCAATAG 389

66 gHisValAspArgAsnGluGlyGlyGluAspValGlyArgPheValValG 83

|||||

390 GCATGTTGATCGCAATGAAGGGGAGAGGATGTTGGGAGATTTGTAGTGC 439

83 lngLysThrGluValLysArgLysThrThrGluGlnGlnValArgProTyr 99

|||||

440 AGGGAACAGAAGTCAAGAAAGAACTACGGAGCAGCAGGTGAGGCCCTTAC 489

100 ArgArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeuI 116

|||||

490 AGGGTTTCCGAACCCGGAACCTGACAAATCATTTACGACTTTTGCCTCAT 539

116 ePro 117

|||||

540 ACCT 543

seq\_name: gb\_est1:AA473525

#### seq\_documentation\_block:

LOCUS AA473525 458 bp mRNA EST 18-JUN-1997

DEFINITION vg78a09.r1 Barstead MPLRB1 Mus musculus cDNA clone IMAGE:872056 5' similar to SW:HG74 HUMAN Q00994 OVARIAN GRANULOSA CELL 13.0 KD

PROTEIN HGR74. ;, mRNA sequence.

ACCESSION AA473525

VERSION AA473525.1 GI:2201752

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 458)

Marta,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

#### TITLE

#### JOURNAL

#### COMMENT





OK of: US-09-327-750D-35 to: EST:\* out\_format : pfs

Date: Mar 11, 2002 2:16 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=framet\_p2n.model -DEV=xlh  
-O/cgnt2\_1/USPTO.spool/US09327750/runat\_11032002\_101153\_20308/app\_query.fasta\_1.1472  
-DB=EST -QPM=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.050 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -EGAPOP=6.000  
-GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09327750 -CGN1\_1\_5654  
-NCPU=6 -TCPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-35

Query length: 117

Database: EST:\*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 4085.940000

score\_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_btc:AK010400	+	611.50	1049.04	2.9e-49	789	AK010400 Mus musculus ES cells
gb_est2:BG083261	+	611.50	1049.02	2.9e-49	792	BG083261 H308C08-5 NIA Mouse 1
gb_est1:AA743525	+	605.00	1041.70	7.4e-46	458	AA743525 vg78A09.r1 Barstead MF
gb_est2:BG070341	+	575.50	987.47	7.7e-46	747	BG070341 H308C08-3 NIA Mouse 1
gb_est1:BF723075	+	568.50	977.99	4.7e-45	519	BF723075 mab27e12.y1 Soares.NME
gb_est1:BE333774	+	544.50	937.49	2.6e-43	463	BE333774 us27h06.y1 Soares.NME
gb_est2:W17712	+	541.00	934.14	7.2e-43	317	W17712 mb77e04.r1 Soares mouse
gb_est1:AI227867	+	539.50	927.44	1.7e-42	568	AI227867 EST24552 Normalized r
gb_est2:BI289946	+	511.50	878.38	9.2e-40	643	BI289946 UI-R-DKO-cff-c-12-0-UI
gb_est1:AA117429	+	503.50	859.71	2.8e-39	312	AA117429 mm21e12.r1 Beddington
gb_est2:W35893	+	484.50	834.09	2.7e-37	472	W35893 mc5309.r1 Soares mouse
gb_est2:BF722027	+	479.00	828.16	5.8e-37	286	BF722027 mab27e12.y1 Soares.NME
gb_est2:W54487	+	469.50	808.38	7.3e-36	465	W54487 md09c11.r1 Soares mouse
gb_est1:AL118340	+	463.50	798.74	2.5e-35	422	AL118340 y911b26 Beddington mc
gb_est1:AW251984	+	459.00	789.25	8.5e-35	540	AW251984 UI-R-Bj0-adn-b-06-0-UI
gb_est2:W83403	+	426.50	734.73	9.2e-32	442	W83403 mf58d02.r1 Soares mouse
gb_est1:AI012535	+	373.00	642.62	1.2e-26	444	AI012535 EST206986 Normalized r
gb_est1:AV137854	+	368.00	636.78	2.6e-26	300	AV137854 AV137854 Mus musculus
gb_est1:AA516739	+	365.00	631.85	5.0e-26	290	AA516739 vh84c04.r1 Knowles Sol
gb_est2:W61422	+	363.50	630.70	5.8e-26	237	W61422 md29g02.r1 Soares mouse
gb_est1:AI517301	+	355.50	609.93	9.2e-25	722	AI517301 AL517301 LTI_NFL011.NE
gb_est2:BE883414	+	355.50	609.07	8.2e-25	739	BE883414 601509167F1 NIH_MGC_71
gb_est1:AL544124	+	355.50	607.79	1.1e-24	845	AL544124 AL544124 LTI_NFL006.PL
gb_est1:AU117899	+	355.50	607.95	1.1e-24	865	AU117899 AU117899 HEMBAI Homo s
gb_est2:BG122197	+	355.50	607.78	1.1e-24	866	BG122197 602349748F1 NIH_MGC_90
gb_est1:AL545120	+	355.50	607.47	1.1e-24	905	AL545120 AL545120 LTI_NFL006.PL
gb_est2:BG709427	+	355.50	607.29	1.2e-24	929	BG709427 602673481F1 NIH_MGC_96
gb_est2:BG035675	+	355.50	606.94	1.2e-24	975	BG035675 602325745F1 NIH_MGC_90
gb_est1:AL560650	+	355.50	606.89	1.2e-24	983	AL560650 AL560650 LTI_NFL010.BC
gb_est1:AL539004	+	355.50	606.61	1.3e-24	1023	AL539004 AL539004 LTI_FLO13.PH
gb_est1:AL549875	+	355.50	606.46	1.3e-24	1044	AL549875 AL549875 LTI_NFL006.PL
gb_est2:BF930215	+	351.00	602.88	2.0e-24	579	BF930215 IL5-NT0227-111200-319
gb_est1:AL534630	+	350.50	602.18	2.2e-24	566	AL534630 AL534630 LTI_FLO13.FBI
gb_est2:AL523320	+	349.00	595.28	5.4e-24	1043	AL523320 AL523320 LTI_NFL003.N
gb_est1:AW385267	+	347.50	597.43	4.1e-24	534	AW385267 RC0-LT0001-261119-011
gb_est2:BG527431	+	347.50	594.82	5.7e-24	772	BG527431 602557227F1 NIH_MGC_59
gb_est2:BG009563	+	346.50	595.11	5.5e-24	581	BG009563 QV1-GN0319-021200-526
gb_est1:AL582060	+	346.00	590.16	1.0e-23	1037	AL582060 AL582060 LTI_NFL010.H
gb_est2:BI040887	+	344.50	592.88	7.4e-24	489	BI040887 QV3-NT0278-120201-502

gb\_est1:AL573904 - 344.50 586.65 1.6e-23 1182  
gb\_est2:BI040889 - 340.50 584.98 2.0e-23 565  
gb\_est1:AW934883 + 338.50 580.35 3.7e-23 668  
gb\_est1:AW934905 + 338.50 580.18 3.8e-23 685  
gb\_est1:AW934769 + 338.00 579.56 4.1e-23 662  
seq\_name: gb\_htc:AK010400

seq\_documentation\_block: 789 bp mRNA HTC 05-JUL-2001  
LOCUS AK010400  
DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library,  
clone:2410004M13, full insert sequence.

ACCESSION AK010400  
VERSION AK010400.1 GI:12845816  
KEYWORDS CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,  
clone\_lib:RIKEN full-length enriched mouse cDNA library  
clone:2410004M13.

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 789)

Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning

Methods in enzymology. 303, 19-44 (1999)

99279253

PUBMED 10349636

2 (bases 1 to 789)

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome research. 10 (10), 1617-1630 (2000)

20499374

3 (bases 1 to 789)

Shibata,K., Itoh,M., Aizawa,K., Nagao,S., Sasaki,N., Carninci,P.,

Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,

Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,

Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,

Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multiplexed sequencer

Genome research. 10 (11), 1757-1771 (2000)

20530913

4 (bases 1 to 789)

Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,

Hanaka,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Horii,F.,

Inotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,

Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,

Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,

Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,

Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takehashi,F.,

Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,

Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,

Fax:81-45-503-9216)

TITLE

JOURNAL

Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM10688 row: p column: 05  
 High quality sequence stop: 712.

## FEATURES

Location/Qualifiers

source

1..734  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4799476"  
 /clone\_lib="NIH\_MGC\_96"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcagag); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTTNN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 187 a 174 c 214 g 159 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 646.00 Length: 125  
 Ratio: 5.252 Gaps: 0  
 Percent Similarity: 98.400 Percent Identity: 97.600

## alignment\_block:

US-09-327-750D-32 x BG715659 ..

Align seg 1/1 to: BG715659 from: 1 to: 734

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
 |||||  
 206 ATGGAGTCCAAAGAGAAACAGCAGTAAACAGTCTCAGCATGGAATAATGC 255  
 |||||  
 17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34  
 |||||  
 256 CAACCAAGAAATGAAGGAAGAGCAGCAAGTTGCTAATAAGGGGAGCCCT 305  
 |||||  
 34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50  
 |||||  
 306 TGGCCCTCCCTTTGGATGCTGGTGTGTAATCTGTGCTGTAGAGGAATCGT 355  
 |||||  
 51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTyrAspIleMe 67  
 |||||  
 356 AGCGCGTTCGCGCTTAGCGAGCCCATCTTCAGTATAGATGGATATGAT 405  
 |||||  
 67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84  
 |||||  
 406 GCATAGCTTGGAGACCAACAGCAGGAGTGAAGAGAGATATGGA 455  
 |||||  
 84 rgileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100  
 |||||  
 456 GGATTTGGGAGGAGGTGAGACACCTGATGGAAAGCTGAGGGAAAGCAG 505  
 |||||  
 101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117  
 |||||  
 506 TTGAGTCATAGTCTGCGGGCAGTCAGCACTGACCCCTCCATCACCATGACCA 555  
 |||||  
 117 sHisAspGluPheCysLeuMetPro 125  
 |||||  
 556 TCATGATGAGTGTGCTTATGCC 580

seq\_name: gb\_est2:BG709325

seq\_documentation\_block:

LOCUS BG709325 828 bp mRNA EST 07-MAY-2001  
 DEFINITION 602673345F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4795882 5',  
 mRNA sequence.

ACCESSION BG709325

VERSION BG709325.1 GI:13987550

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 828)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM10679 row: j column: 11

High quality sequence stop: 798.

## FEATURES

Location/Qualifiers

source

1..828

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4795882"

/clone\_lib="NIH\_MGC\_96"

/tissue\_type="hypothalamus"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcagag); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTNN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT

ORIGIN

232 a 179 c 228 g 189 t

## alignment\_scores:

Quality: 637.00 Length: 126  
 Ratio: 5.137 Gaps: 1  
 Percent Similarity: 98.413 Percent Identity: 96.825

## alignment\_block:

US-09-327-750D-32 x BG709325 ..

Align seg 1/1 to: BG709325 from: 1 to: 828

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
 |||||  
 176 ATGGAGTCCAAAGAGAAACAGCAGTAAACAGTCTCAGCATGGAATAATGC 225  
 |||||  
 17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34  
 |||||  
 226 CAACCAAGAAATGAAGAAAGGAGCAAGTTGCTAATAAGGGGAGCCCT 275  
 |||||  
 34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50  
 |||||  
 276 TGGCCCTCCCTTTGGATGCTGTGTAATCTGTGCTGTAGAGGAATCGT 325  
 |||||





```
7111 CTAGAGGCCCGCCAGAAAAGAA.....AAGCAAAATACCACCTAAAGTCTTT 7154
17 aasnGlnGluasnGluGlnValAlaAsnLysGlyGluProL 34
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7155 AGAGAAAGAAAGTGTAGAGTCAGAGAAACTCTAGAAAATGGGGTCTCG 7204
34 euAlaLeuProLeu.....AspAlaGlyGlu 42
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7205 TACTATCTGAGCTTTAGGAAAGAACACACAGACAGAGGATCAAGAA 7254
43 TyrCysValProArgGlyAsnArgArgPhe..... 53
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7255 TTAATGCTCTCTAAAGGTACACTAAAGAGATTTTCATCTCTAGGAAAGGA 7304
54 .....ProValArgGlnProIleLeuGln..... 61
7305 AAGTCAAGAAGTAGTGAGGCTTCAAAGAGAGGGAACCTAGAAATCATGGA 7354
62 .....TyrArgTTPaspIleMetHisArgLeuGlyGluProGlnAlaArg 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7355 CAGCTTTTAAAGAGGAGAGCCACACCCACTGGGATTTCCAGGAGTGTAG 7404
76 ..... 76
7405 GACCAGATGCTTGAGAGACTGGTAGAGAAAGAGGATCAGAGCTTCCCAAG 7454
77 .....MetArgGluGluA 81
7455 GTCTCAGAGGAAGAGAGCAGGAGGCATGTAGACCTCTGCAGAAAGAGA 7504
81 snMetGluArgIleGly.....GluGluValArgGlnLeuMetGluLys 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7505 ATCAGGAACCATAGGTATGAGAGAGCAGAGCGGCCAGATACTTGTAGAGA 7554
96 LeuArgGluLysGlnLeuSerHisSerLeuArgAla 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7555 CTGATAGAAAAGAGAGTCAAGAGTCCCTGAGGTCT 7590

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-306-691B-23
```

```
seq_documentation_block:
; Sequence 23, Application US/08306691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavorana & Monaco, P.C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306.691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Moraco Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5734039e
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2301 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-306-691B-23

alignment_scores:
   Quality: 84.00      Length: 102
   Ratio: 1.355      Gaps: 4
   Percent Similarity: 60.784      Percent Identity: 32.353

alignment_block:
US-09-327-750D-32 x US-08-306-691B-23 ..
Align seg 1/1 to: US-08-306-691B-23 from: 1 to: 2301

2 GluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAs 18
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
77 GAGGAGCAGGAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 126
18 nGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProLeuA 35
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCTG 176
35 laLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArgArg 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
177 CAAGCCGAG.....CGGAGGAGG 195
52 ArgPheProValArgGlnProIleLeu.....GlnTyrArgTrpAspI 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
196 CAGGAACCGGAGCGGAGGAGGAGTAGCTGGGTGGGACCACTGGCTGGCATCA 245
66 eMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMet 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
246 CCACCATCGAGCGGTGAAGCGCAAGATCCAGGTTCGCGAGCAGCAGGCA 295
83 GluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluL 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
296 GATGATGCAGAGGAGCGGAGCTCAGCGCCTCCAGCCGAGAACTTGAGGGAGA 345
99 ys 99
   : :
346 AA 347

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-06251-78

seq_documentation_block:
; Sequence 78, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
```

```

; FILING DATE: 19930630
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2301 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-78

alignment_scores:
  Quality: 84.00      Length: 102
  Ratio: 1.355        Gaps: 4
Percent Similarity: 60.784 Percent Identity: 32.353

alignment_block:
US-09-327-750D-32 x PCT-US93-06251-78 ..

Align seg 1/1 to: PCT-US93-06251-78 from: 1 to: 2301

2 GluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAs 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 GAGGAGCAGGAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 126

18 nGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProLeuA 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
127 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 176

35 laLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArgArg 51
|| |||
177 CACGCCGAG.....CGGAGGAGG 195

52 ArgPheProValArgGlnProLeu.....GlnTyrArgTrpAspTrl 66
::: ||| |||:|||||:|||||:|||||:|||||:|||||:|||||
196 CAGGAACCGGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 245

66 eMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMet 82
||||| |||||| ||| |||:|||||:|||||:|||||:|||||:|||||
246 CCACCATCGAGCGGCGTGAACCGCACAGATCCAGGTCTGCAGCAGCAG 295

83 GluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluL 99
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
296 GATGATCCAGAGGAGCGAGCTGAGCGCCCTCCAGCGAGAAGTTGAGG 345

99 ys 99
||
346 AA 347

```

```
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-872-644-52
```

```

seq_documentation_block:
; Sequence 52, Application US/07872644
; Patent No. 5389527
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
; TITLE OF INVENTION: Phosphodiesterases
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

```

```

; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,644
; FILING DATE: 19920420
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/688,356
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5389527and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2077 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1693
US-07-872-644-52

alignment_scores:
  Quality: 80.50      Length: 126
  Ratio: 1.118        Gaps: 6
Percent Similarity: 57.143 Percent Identity: 24.603

alignment_block:
US-09-327-750D-32 x US-07-872-644-52 ..

Align seg 1/1 to: US-07-872-644-52 from: 1 to: 2077

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
959 CAGAGCGGTTGAGTTTGAATAGCATCAGCTCGTCAGATGCCAAGCGATC 1008

20 :
;
1009 AGGTGTCAAGACCTCTGTTTCAGAGGGAAGTCCCGCATCAACAATCTG 1058

31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1059 TCATCTCCGTTGACTATAAGAGCTTTAAAGCTACTTTGGACGGAAGTGTG 1108

44 sval.....ProArgGlyAsnArgArgArgPheP 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1109 CACATCAATCGGAGAGATGGAGGCGCAAGGTACCCAAAGAGGAGGAGGC 1158

54 roValArgGlnProIleLeuGlnTyrArgTrpAsp.....IleMetHis 68
|| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1159 CAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1208

69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI 85
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
1209 AGGAAATGGAAGCCAAAGCCAGGCTGAAGAAGGCGCATCTGCGCAAAGCT 1258

```

4 LysGluLysArgAlaValAsnSerLeuSerMetClnAsnAlaAsnGlnI 20  
959 CAGAGCGCTTCGAGTTTGAATAGCATGCAGTCGTCCAGTCCCAAGGCATC 1008  
20 u.....AsnGluGluLysGluGluInValAlaAsnLys. 30  
1009 AGGCTCAAGACCTTCGCTTCAGAGGCAACTGCCCCCACAACAAATCTTC 1058

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-297-494-52

```

1059 TCATCTCCGTTTCACTACTAAGAGCTTTAAAGCTTACTTGGACGGAAGTGGTG 1108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
44  sval.....ProArgGlyAsnArgAtgArgPhep 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1109 CACATCAATCGGAGAGATGAGAGCCCAAGTACCAAGAGGAGAAGGC 1158
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
54  rovalArgGlnProIleLeuGlnTyrArgTTrpasp.....IleMethis 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1159 CAAGAAGCAAGACAGAGGAAAAAGGCTCGCTGGCCGACAGAGGACGACAAA 1208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
69  ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgIl 85
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1209 AGGAANTGGAAGCCCAAAAGCCAGGCTGAAGAAGGCGCATCTGGCAAGCT 1258
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
85  eGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeus 102
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1259 GAGA.....AAGACAGCTCTGGAGAACTAAGAATCAAGTCAATG 1299
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
102  erHisSerLeuArgAlaValSerThr 110
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1300 AACACGGGCAAAACAAAGTGCACAACC 1325
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-297-510-52

seq_documentation_block:
: Sequence 52, Application US/08297510
: Patent No. 5602019
: GENERAL INFORMATION:
: APPLICANT: Beavo, Joseph A.
: APPLICANT: Bentley, Kelley
: APPLICANT: Charbonneau, Harry
: APPLICANT: Sonnenburg, William K.
: TITLE OF INVENTION: DNA Encoding Mammalian
: TITLE OF INVENTION: Phosphodiesterases
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
: ADDRESSEE: Bicknell
: STREET: Two First National Plaza, 20 South Clark
: STREET: Street
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/297 510

```

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/688,356  
FILING DATE: 04-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: NO. 5602019and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/30822  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 346-5750



```
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1693
; US-08-297-510-52

alignment_scores:
    Quality: 80.50      Length: 126
    Ratio: 1.118      Gaps: 6
Percent Similarity: 57.143      Percent Identity: 24.603

alignment_block:
US-09-327-750D-32 x US-08-297-510-52

Align seg 1/1 to: US-08-297-510-52 from: 1 to: 2077

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG1 20
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
959 CAGAGCGTTCGAGTTGAATAGCATCAGCTCGTCAGATGCCAAGGCATC 1008
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
20 u.....AsnGluGluLysGluGlnValAlaAsnLys. 30
1009 AGGTGTCAAGACCTCTGGTTTCAGAGGGAAGTCCCGCATCAACAATTCG 1058
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1059 TCATCTCCGTTGACTATAAGAGCTTTAAAGCTACTTTGGACGGAAGTGGTG 1108
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
44 sVal.....ProArgGlyAsnArgArgPheP 54
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1109 CACATCAATCGGAGAGATGGAGGCCAAGGTACCCAAAGAGAGAGAGGC 1158
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
54 roValArgGlnProIleLeuGlnTyrArgTrpAsp.....IleMetHis 68
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1159 CAAGAAGGAGCAGAGAGAAAGGCTCGCTGCGCGCAGAGAGAGAGAGGC 1208
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI1 85
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1209 AGGAATGGAAGCCAAAGCCAGGCTGAAGAAGGCGCATCTGGCAAGCT 1258
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
85 eGlyGluGluValArgGlnLeuMetGluLysLeuArgGluGluLysGlnLeuS 102
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1259 GAGA.....AAAGACGCTCTGGAGAACTAGAACTCAAGTCAATGG 1299
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
102 erHisSerLeuArgAlaValSerThr 110
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1300 AACACGCGCAACAAAGTGACAAC 1325
: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-479-532-52

seq_documentation_block:
; Sequence 52, Application US/08479532
; Patent No. 5776752
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
; TITLE OF INVENTION: Phosphodiesterases
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
```



```

: TITLE OF INVENTION: Phosphodiesterases
:
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
: ADDRESSEE: Bicknell
: STREET: Two First National Plaza, 20 South Clark
: STREET: Street
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/139,491
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/455,525
: FILING DATE: 31-MAY-1995
: APPLICATION NUMBER: 08/297,494
: FILING DATE:
: APPLICATION NUMBER: US 07/688,356
: FILING DATE: 04-APR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 6015677and, Greta E.
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 27866/30822
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 346-5750
: TELEFAX: (312) 984-9740
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 52:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2077 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..1693
: US-09-139-491-52
:
: alignment_scores:
: Quality: 80.50 Length: 126
: Ratio: 1.118 Gaps: 6
: Percent Similarity: 57.143 Percent Identity: 24.603
:
: alignment_block:
: US-09-327-750D-32 x US-09-139-491-52 ..
:
: Align seg 1/1 to: US-09-139-491-52 from: 1 to: 2077
:
: 4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnGI 20
: : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : :
: 959 CAGAGGCGTTCGAGTTGAATGATCAGTCAGTCGTCAGATGCCAAGCATC 1008
:
: 20 u.....AsnGluLysGluGlnValAlaAsnLys. 30
: : : : : : : : : : : : : : : : : : : : : : : : : : :
: 1009 AGGTGTCAGACCTCTGGTTCAGAGGGAAGTGCCCGCATCAACAATTCTG 1058
:
: 31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
: : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : :
: 1059 TCATCTCCGTTGACTATAGAGCTTTAAAGCTACTTGGCGGAAGTGGTG 1108
:
: 44 sVal.....ProArgGlyAsnArgArgPheP 54
: : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : :
: 1109 CACATCAATCGGAGAGATGGAGGGCGCAAGGTACCCAAAGAGGAGGAGGC 1158

```

```

54 roValArgGlnProIleLeuGlnTyrArgTrpAsp.....IleMetHis 68
|||:::
1159 CAAGAAGGAGCAGAGGAAAGCTCGCTGGCCGCGAGAGGACGACAA 1208
|||:::
69 ArgLeuGlyCluProGlnAlaArgMetArgGluGluAsnMetGluArgII 85
|||:::
1209 AGGAAATGGAAGCAAAAGCCAGGCTGAAGAGGCGCATCTGGCAAAGCT 1258
|||:::
85 eGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeuS 102
:|:::
1259 GAGA.....AAAGAGCTGTGGAGAACTAAGAATCAAGTCAATGG 1299
:|:::
102 erHisSerLeuArgAlaValSerThr 110
|||:::
1300 AACACGGGCAAAACAAAGTGACAACC 1325
|||:::

```

seq\_name: /cgn2\_6/ptodata/2/1na/PCTUS\_COMB.seq:PCT-US92-03222-52

```

seq_documentation_block:
: Sequence 52, Application PC/TUS9203222
: GENERAL INFORMATION:
: APPLICANT: Beavo, Joseph A.
: APPLICANT: Bentley, Kelley
: APPLICANT: Charbonneau, Harry
: APPLICANT: Sonnenburg, William K.
: TITLE OF INVENTION: DNA Encoding Mammalian
: TITLE OF INVENTION: Phosphodiesterases
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
: ADDRESSEE: Bicknell
: STREET: Two First National Plaza, 20 South Clark
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/03222
: FILING DATE: 19920420
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/688,356
: FILING DATE: 04-APR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Noland, Greta E.
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 27866/30822
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 346-5750
: TELEFAX: (312) 984-9740
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 52:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2077 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..1693
PCT-US92-03222-52

```

alignment\_scores: Quality: 80.50 Length: 126

```

Ratio: 1.118 Gaps: 6
Percent Similarity: 57.143 Percent Identity: 24.603
alignment_block:
US-09-327-750d-32 x PCT-US92-03222-52
Align seg 1/1 to: PCT-US92-03222-52 from: 1 to: 2077
4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG1 20
959 CAGAGCGGTTTCGAGTTTGAATAGCATCAGCTCGTCAGATGCCAAGGATC 1008
:|:::
20 u.....AsnGluGluLysGluGlnValAlaAsnLys. 30
:|:::
1009 AGGTGTCAAGACCTCTGGTTTCAGAGGGAAGTGCCTCCGATCAACAATCTG 1058
:|:::
31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
:|:::
1059 TCATCTCCGTTGACTATAGAGAGCTTTAAAGCTACTTGGACGGAAGTCGTG 1108
:|:::
44 sVal.....ProArgGlyAsnArgArgPhep 54
:|:::
1109 CACATCAATCGGAGAGATGGAGGCCAAGGTACCCAAAGAGGAGGAAGGC 1158
:|:::
54 roValArgGlnProIleLeuGlnTyrArgTrpAsp.....IleMetHis 68
|||:::
1159 CAAGAAGGAGCAGAGGAAAGCTCGCTGGCCGCGAGAGGACGACAA 1208
|||:::
69 ArgLeuGlyCluProGlnAlaArgMetArgGluGluAsnMetGluArgII 85
|||:::
1209 AGGAAATGGAAGCAAAAGCCAGGCTGAAGAGGCGCATCTGGCAAAGCT 1258
|||:::
85 eGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeuS 102
:|:::
1259 GAGA.....AAAGAGCTGTGGAGAACTAAGAATCAAGTCAATGG 1299
:|:::
102 erHisSerLeuArgAlaValSerThr 110
|||:::
1300 AACACGGGCAAAACAAAGTGACAACC 1325
|||:::

```

seq\_name: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:US-07-872-644-50

```

seq_documentation_block:
: Sequence 50, Application US/07872644
: Patent No. 5389527
: GENERAL INFORMATION:
: APPLICANT: Beavo, Joseph A.
: APPLICANT: Bentley, Kelley
: APPLICANT: Charbonneau, Harry
: APPLICANT: Sonnenburg, William K.
: TITLE OF INVENTION: DNA Encoding Mammalian
: TITLE OF INVENTION: Phosphodiesterases
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
: ADDRESSEE: Bicknell
: STREET: Two First National Plaza, 20 South Clark
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/872,644
: FILING DATE: 19920420
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/688,356

```

```
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5389527and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2693 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 176..2077
; US-07-872-644-50

alignment_scores:
  Quality: 80.50      Length: 126
  Ratio: 1.118       Gaps: 6
  Percent Similarity: 57.143   Percent Identity: 24.603

alignment_block:
US-09-327-750D-32 x US-07-872-644-50 ..

Align seg 1/1 to: US-07-872-644-50 from: 1 to: 2693

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnG1 20
1568 CAGAGCGGTCGAGTTGTAATGACATCAGCTCGTCAGATGCCAAGCGATC 1617

20 u.....AsnGluGluLysGluGlnValAlaAsnLys. 30
1618 AGGTGTCAAGACCTCTGTTTCAGAGGGAAGTCCCGCATCAACAATTCG 1667

31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTyrCy 44
1668 TCATCTCCGTTGACTATAAGAGCTTTAAAGCTACTTGGACGGAAGTGGTG 1717

44 sVal.....ProArgGlyAsnArgArgPhep 54
1718 CACATCAATCGGAGAGATGGAGGCGCAAGTACCCAGAGAGAGAGCG 1767

54 roValArgGlnProLeuGlnTyrArgTrpAsp.....IleMetHis 68
1768 CAAGAAGGAAGCAGAGAGGAAGAGGCTCGCTCGCGCAGAGGAGCAGCAA 1817

69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI1 85
1818 AGGAATGAAGCCAAAGCCAGGCTTGAAGAGGCGCATCTCGCAAGCT 1867

85 eGlyGluGluValArgGlnGluMetGluLysLeuArgGluLysGlnLeus 102
1868 GAGA.....AAAGACGCTCTGGAGAACTAAGATCAAGTCAATGG 1908

102 erHisSerLeuArgAlaValSerThr 110
1909 AACACGGGCAACAAAGTGACAACC 1934

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-297-494-50

seq_documentation_block:
; Sequence 50, Application US/08297494
; Patent No. 5580771
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
```

[illegible]

31 GlyGluProLeu.....AlaLeuProLeuAspAlaGlyGluTrpCy 44  
: : : : :  
: : : : :  
: : : : :  
1668 TCATCTCCGTTGACTATAAGACGCTTAAAGCTACTTTGGACGGAAGTCGTG 1717

44 sVal.....ProArgGlyAsnArgArgArgPhep 54  
l : : : : :  
l : : : : :  
l : : : : :  
1718 CACATCAATCGGGAGAGATGGAGGCCAAGGTACCCTAACAGAGAGAAGGC 1767

54 rovalArgGlnProIleLeuGlnTrpArgTrpAsp.....IleMetHis 68  
||| ||||: : : : :  
||| ||||: : : : :  
1768 CAAGAAGGGAAGCAGAGGAAAAGGCTCGCTGGCCGAGAGGACGACCAA 1817

69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgII 85  
||| |||: : : : :  
||| |||: : : : :  
1818 AGGAANTCGAAGCCCAAAAGCCAGGCTGAAGAAGGCGCATCTGGCAAGACT 1867

85 eGlyCluGluValArgGlnLeuMetGlnLysLeuArgGluLysGlnLeu 102  
: : : : :  
: : : : :  
: : : : :  
1868 GAGA.....AAAGACGCTCTGGAGAACTAAGAATCAAGTCAATGG 1908

102 erHisSerLeuArgAlaValSerThr II10  
: : : : :  
: : : : :  
: : : : :  
1909 AACACGGGCAACAACAAAGTGACAACC 1934

[illegible][illegible]

1868 GAGA.....AAAGACGTCTGGAGAACTAAGAATCAAGTCAATGG 1908  
102 erHisSerLeuArgAlaValSerThr 110  
1909 AACACGGGCAACAAACAAAGTGAACGC 1934

```

102  erHisSerLeuArgAlaValSerThr 110
      :::::  :::::
109  AACACGGGGCAACAAAGTGACAACC 1934

```

1909 AACACGGGCAAAACAAAAGTGACAACC 1934



**THIS PAGE BLANK (USPTO)**



OM of: US-09-327-750D-32 to: N\_Geneseq\_1101.\* out\_format : pfs  
Date: Mar 11, 2002 3:42 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101154\_20362/app\_query.fasta\_1.1472  
-DB=N\_Geneseq\_1101 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HRAFSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09327750.@CNI\_1\_330 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-327-750D-32  
Query length: 125  
Database: N\_Geneseq\_1101.\*  
Database sequences: 930621  
Database length: 428662619  
Search time (sec): 355.560000

score\_list:

Sequence	Strd Orig	zScore	EScore Len	Documentation
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH45143 +	639.00	1290.82	792	8.0e-64
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH75810 +	602.50	1216.52	767	1.1e-59
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH58581 +	602.50	1215.36	862	1.3e-59
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH59611 +	602.50	1214.95	898	1.3e-59
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH03067 +	589.50	1188.82	858	3.8e-58
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH03880 +	585.50	1183.24	7	7.9e-58
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH03879 +	585.50	1182.71	698	8.4e-58
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH01089 +	562.50	1135.77	692	3.5e-55
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH06183 +	207.50	414.96	421	4.9e-15
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH06180 +	207.50	414.14	457	5.4e-15
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH03515 +	202.50	397.55	865	4.6e-14
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH13750 +	202.50	394.04	1229	7.2e-14
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH83548 +	202.50	393.00	1364	8.2e-14
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH23529 +	201.00	394.18	891	7.0e-14
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH23528 +	185.50	362.21	917	4.2e-12
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH23528 +	185.50	363.88	700	3.4e-12
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH06184 +	184.50	367.25	451	2.2e-12
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH01005 +	151.50	298.13	532	1.6e-08
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH01004 +	136.50	268.39	485	7.1e-07
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH06181 +	117.00	234.04	279	5.8e-05
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH01003 +	106.00	207.93	401	0.0017
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH87147 +	100.50	204.30	187	0.0026
/SIDS2/qcgdata/geneseq/geneseq/NA1998.DAT:AAH14053 +	89.00	160.28	1459	0.7498
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH82586 +	88.50	171.55	426	0.1767
/SIDS2/qcgdata/geneseq/geneseq/NA1992.DAT:AAH28398 +	86.50	141.15	5946	8.73
/SIDS2/qcgdata/geneseq/geneseq/NA1994.DAT:AAH70447 +	86.50	134.79	11236	19.71
/SIDS2/qcgdata/geneseq/geneseq/NA1998.DAT:AAV20445 +	84.00	145.51	2301	4.99
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH09303 +	84.00	145.51	2301	4.99
/SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH15730 +	81.00	135.10	3532	18.95
/SIDS2/qcgdata/geneseq/geneseq/NA1992.DAT:AAH30188 +	80.50	139.38	2077	10.95
/SIDS2/qcgdata/geneseq/geneseq/NA1995.DAT:AAH83982 +	80.50	139.38	2077	10.95
/SIDS2/qcgdata/geneseq/geneseq/NA1997.DAT:AAH67203 +	80.50	139.38	2077	10.95
/SIDS2/qcgdata/geneseq/geneseq/NA1997.DAT:AAH51117 +	80.50	139.38	2077	10.95
/SIDS2/qcgdata/geneseq/geneseq/NA1998.DAT:AAV4766 +	80.50	139.38	2077	10.95
/SIDS2/qcgdata/geneseq/geneseq/NA1998.DAT:AAV48209 +	80.50	139.38	2077	10.95
/SIDS2/qcgdata/geneseq/geneseq/NA1998.DAT:AAV36161 +	80.50	139.38	2077	10.95
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAH290385 +	80.50	136.79	2693	15.27
/SIDS2/qcgdata/geneseq/geneseq/NA1995.DAT:AAH83981 +	80.50	136.79	2693	15.27
/SIDS2/qcgdata/geneseq/geneseq/NA1997.DAT:AAH67202 +	80.50	136.79	2693	15.27
/SIDS2/qcgdata/geneseq/geneseq/NA1997.DAT:AAH51116 +	80.50	136.79	2693	15.27

/SIDS2/qcgdata/geneseq/geneseq/NA1998.DAT:AAV54765 +	80.50	136.79	15.27	2693
/SIDS2/qcgdata/geneseq/geneseq/NA1998.DAT:AAV48208 +	80.50	136.79	15.27	2693
/SIDS2/qcgdata/geneseq/geneseq/NA1998.DAT:AAV36160 +	80.50	136.79	15.27	2693
/SIDS2/qcgdata/geneseq/geneseq/NA2000.DAT:AAZ90384 +	80.50	136.79	15.27	2693
/SIDS2/qcgdata/geneseq/geneseq/NA1999.DAT:AAZ20086 +	80.50	124.90	70.07	8854
seq_name: /SIDS2/qcgdata/geneseq/geneseq/NA2001.DAT:AAH45143				
seq_documentation_block:				
ID	AAH45143	standard; cDNA; 792 BP.		
XX	AAH45143;			
AC	AAH45143;			
DT	07-SEP-2001	(first entry)		
XX	Human brain expressed X-linked protein, hbex, coding sequence.			
DE	Human; brain expressed X-linked protein; cytostatic; auditory; nontropic; hbex; dysembryoplasia; hereditary disease; cancer; tumour; deafness; X-chromosome-binding mental retardation; lissencephalous disease; ss.			
XX	Homo sapiens.			
XX	WO200140286-A1.			
PN	07-JUN-2001.			
PD	27-NOV-2000; 2000WO-CN00502.			
XX	30-NOV-1999; 99CN-0124179.			
PR	(BIOR-) BIOROAD GENE DEV LTD SHANGHAI.			
XX	Mao Y, Xie Y;			
PI	WPI; 2001-397944/42.			
XX	P-PSDB; AAB99224.			
DR	Isolated human brain-expressed X-linked polypeptide used to diagnose and treat of dysembryoplasia, hereditary diseases, cancer, tumor, deafness and X-chromosome-binding mental retardation			
PT	Claim 5; Page 22; 30pp; Chinese.			
XX	The present sequence is the coding sequence for a human brain-expressed X-linked protein (hbex). hbex and its coding sequence are useful in the diagnosis and treatment of dysembryoplasia, hereditary diseases, cancer, tumours, deafness, X-chromosome-binding mental retardation and lissencephalous disease. hbex is also useful for screening mimics, agonists, or inhibitors, and in peptide fingerprinting identification.			
CC	hbex coding sequence can be used as primers or probes, or in producing gene chips or microarrays.			
CC	Sequence 792 BP; 214 A; 172 C; 219 G; 187 T; 0 other;			
SQ	alignment_scores:			
	Quality: 639.00	Length: 125		
	Ratio: 5.238	Gaps: 0		
	Percent Similarity: 97.600	Percent Identity: 96.800		
alignment_block:				
US-09-327-750D-32 x AAH45143				
Align seg 1/1 to: AAH45143 from: 1 to: 792				
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17				
173 ATGGAGTCCNAGAGAACTAGCAGTAAACAGTCTCAGCATGGAATGC 222				
17 aAsnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34				
223 CAACCAAGAAATGAAGAAAAGGAGCAAGTTGCTAATAAAGGGAGGCCT 272				

34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50  
|||||  
273 TGCCCTCCCTTTGGATGCTGGTGAATCTGTGTGCTAGAGGAATCGT 322  
  
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67  
|||||  
323 AGCGGTTCCGCGTTAGGAGCCCATCTGCAGTAGATGGATATGAT 372  
  
67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84  
|||||  
373 GCATAGGCTTGGAGAACACAGGCAAGGATGAGAGAGAGAAATATGGAA 422  
  
84 rgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100  
|||||  
423 GGATGGGAGGGGTGAGACAGCTGATGGAAGAAGCTGAGGGAAGACAG 472  
  
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117  
|||||  
473 TTGAGTCATAGTCTGGGGCAGTCAGCACTGACCCCTCACCATGACCA 522  
  
117 sHisAspGluPheCysLeuMetPro 125  
|||||  
523 TCATGATGAGTTTGGCTTATGCC 547

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH75810

seq\_documentation\_block:  
ID AAH75810 standard; cdna; 767 BP.  
XX  
AC AAH75810;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Human X chromosome linked gene expression protein 14 coding sequence.  
XX  
KW Human; X chromosome linked gene expression protein 14; cancer;  
KW HIV infection; cytostatic; anti-HIV; chromosome X; ss.  
XX  
OS Homo sapiens.  
XX  
PN CN1296969-A.  
XX  
PD 30-MAY-2001.  
XX  
PF 23-NOV-1999; 99CN-0124078.  
XX  
PR 23-NOV-1999; 99CN-0124078.  
XX  
PA (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.  
XX  
PI Mao Y, Xie Y;  
XX  
DR WPI; 2001-483897/53.  
DR P-PSDB; AAG66407.  
XX  
PT Polypeptide-human X chromosome linked gene expression protein 14 and  
PT polynucleotide for coding said polypeptide -  
XX  
PS Claim 6; Page 24 (Disclosure); 31pp; Chinese.  
XX

The present sequence is the coding sequence for human X chromosome linked  
gene expression protein 14. The protein and coding sequence are useful  
for treating diseases e.g. cancer and HIV infection.

Sequence 767 BP; 224 A; 143 C; 215 G; 185 T; 0 other;

alignment\_scores:  
Quality: 602.50 Length: 128  
Ratio: 4.898 Gaps: 1  
Percent Similarity: 96.094 Percent Identity: 89.062

alignment\_block:

US-09-327-750D-32 x AAH75810 ..

Align seg 1/1 to: AAH75810 from: 1 to: 767

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
|||||  
101 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCTGGAAAATGT 150  
  
17 asnGlnGluAsn.....GluLylusGluGlnValAlaAsnLysG 31  
:|||||  
151 CAACACAGGAAATGATGAAANAGATGAAAGGAGCAAGTTGCTTAATAAG 200  
  
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47  
|||||  
201 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTAATACACTGTGTGCCCTAGA 250  
  
48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64  
|||||  
251 GGAAACCGTAGCGGTTCCGCGTTAGGAGCCCATCTGCAGTATAGATG 300  
  
64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81  
|||||  
301 GGACATAATGTCATAGGCTTGGAGAGCCACAGGCAAGGATGAGAGAGAGA 350  
  
81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97  
|||||  
351 ATATGGAAAGGATTTGGGAGGAGGTGAGACAGCTGATGGAAGGCTGAGG 400  
  
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProHis 114  
|||||  
401 GAAAGCAGCTTGAGTCATAGTCTGGGGCAGTCAGCACTGATCCCTCA 450  
  
114 sHisAspHisHisAspGluPheCysLeuMetPro 125  
|||||  
451 CCATGACCATCAGTACGAGTTTGGCTTATGCC 484

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI58581

seq\_documentation\_block:

ID AAI58581 standard; cdna; 862 BP.

XX  
AC AAI58581;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 784.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-brager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-052317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AT, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX WPI: 2001-442253/47.  
DR P-PSDB; AAM39425.  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX Claim 1: SEQ ID NO 784; 10078pp; English.  
PS The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic, and  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 862 BP; 235 A; 183 C; 238 G; 206 T; 0 other;

alignment\_scores:  
Quality: 602.50 Length: 128  
Ratio: 4.898 Gaps: 1  
Percent Similarity: 96.094 Percent Identity: 89.062  
alignment\_block:  
US-09-327-750D-32 x AAI58581 ..

Align seg 1/1 to: AAI58581 from: 1 to: 862

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
|||||  
226 ATGGAGTCCAAAGAGACGACGCGTTAAACAATCTCATCGTGGAAATGT 275  
17 AsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31  
:|||||  
276 CAACAGGAAATGATGATAAAGATGAAAGGACGCAAGTTGCTAATAAG 325  
|||||  
31 lYGluproLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47  
|||||  
326 GGGACCCCTGGCCCTACCTTTGAATGTAGTGAATACTGTGTGCTAGA 375  
|||||  
48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64  
|||||  
376 GGAACCGTAGCGGTTCCGGTTAGGACGCCATCTCTGCAGTAGATG 425  
|||||  
64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81  
|||||  
426 GCACATAATGCATAGCTGGAGAGCCACAGCAGGAGGATGACAGAGGAGA 475  
|||||  
81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97  
|||||  
476 ATATGGAAGAGTGGGAGGAGGTGACACAGCTGATGGAAGCTGAGG 525  
|||||  
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114  
|||||  
526 GAAAGCAGTTGAGTCATAGTTTGGGGCAGTCAGCAGTGTGCTCCCTCA 575  
|||||  
114 shIsAspHisAspGluPheCysLeuMetPro 125  
|||||

576 CCATGACCATCAGCATGAGTTTTTGCCTTATGCC 609  
seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF59611  
seq\_documentation\_block:  
ID AAF59611 standard; cDNA; 898 BP.  
XX  
AC AAF59611;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Human cell cycle and proliferation protein CCYPR-22 cDNA, SEQ ID NO:76.  
XX  
KW Cell cycle and proliferation protein; CCYPR; human; agonist;  
KW antagonist; gene therapy; detection; gene therapy;  
KW transgenic animal disease model; immune disorder;  
KW developmental disorder; cell signalling disorder;  
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
KW arteriosclerosis; asthma; allergy; diabetes mellitus;  
KW menstrual cycle disorder; bacterial infection; ss.  
XX  
OS Homo sapiens.  
XX  
WO200107471-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 21-JUL-2000; 2000WO-US19948.  
XX  
PR 21-JUL-1999; 99US-0145075.  
PR 08-SEP-1999; 99US-0153129.  
PR 10-NOV-1999; 99US-0164647.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;  
PI Azimzai V, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;  
XX  
DR WPI: 2001-112727/12.  
DR P-PSDB; AAB60474.  
XX  
PT Human cell cycle and proliferation proteins and polynucleotides are  
PT used to treat, diagnose and prevent immune, developmental and cell  
PT signaling disorders and cell proliferative disorders including cancer -  
XX  
PS Claim 5: Page 181-182; 205pp; English.  
XX  
CC Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human  
CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.  
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
CC associated with decreased expression of functional CCYPR, while CCYPR  
CC antagonists are used to treat diseases or conditions associated with  
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
CC that specifically bind to CCYPR, and in drug screening methods to  
CC identify compounds that modulate the activity of CCYPR. CCYPR  
CC nucleotides can be used to generate transgenic animal models of human  
CC disease, and can be used in gene therapy in target cells with genetic  
CC abnormalities with respect to the expression of CCYPR for the  
CC treatment or prevention of a disorder associated with CCYPR.  
CC Diseases which can be diagnosed, treated and prevented using CCYPR  
CC proteins, nucleic acids, agonists or antagonists include immune,  
CC developmental and cell signalling disorders, and cell proliferative  
CC disorders including cancer. Specific examples of these disorders  
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
CC diabetes mellitus, disorders of the menstrual cycle and infections  
CC caused by bacteria.  
XX  
SQ Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;

```
alignment_scores:
  Quality: 602.50      Length: 128
  Ratio: 4.898         Gaps: 1
  Percent Similarity: 96.094  Percent Identity: 89.062

alignment_block:
US-09-327-750D-32 x AAF59611 ..
Align seg 1/1 to: AAF59611 from: 1 to: 898

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
254 ATGGAGTCCAAAGAGAGGAGCGGTTAAACAATCTCATCGTGGAAAAATGT 303
17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
304 CAACCCAGGAAATGATGAAAAGAGATGAAAAGGAGCAAGTTGCTAATAAG 353
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
354 GGGAGGCCCTTGGCCCTACCTTTGAATGTTAGTGAATGCTGTGCTAGAG 403
48 GlyAsnArgArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64
404 GGAACCGTGGAGCGGTTCCGCTTAGGCAGCCCATCTGCAGTATAGATG 453
64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81
454 GGACATAATGTCATAGCTTGGAGAGCCACAGCAAGGATGAGAGAGAGA 503
81 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97
504 ATATGAAAAGGATTGGGAGGAGGTGAGACAGCTGATGAGAAAAGCTGAGG 553
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114
554 GAAAGCAGTTGAGTCATAGTCTGGGGGAGTCAGACACTGATCCCCCTCA 603
114 shisAspHisHisAspGluPheCysLeuMetPro 125
604 CCATGACCATCAGATGAGTTTGGCTTATGCCC 637
```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI60367

```
seq_documentation_block:
ID AAI60367 standard; cdna; 858 BP.
XX
AC AAI60367;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4356.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0398042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
```

```
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41211.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 4356; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM4213).with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 858 BP; 236 A; 175 C; 241 G; 206 T; 0 other;
```

```
alignment_scores:
  Quality: 589.50      Length: 129
  Ratio: 4.793         Gaps: 2
  Percent Similarity: 95.349  Percent Identity: 88.372

alignment_block:
US-09-327-750D-32 x AAI60367 ..
Align seg 1/1 to: AAI60367 from: 1 to: 858

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
215 ATGGAGTCCAAAGAGAGGAGCGGTTAAACAATCTCATCGTGGAAAAATGT 264
17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
265 CAACCCAGGAAATGATGAAAAGAGATGAAAAGGAGCAAGTTGCTAATAAG 314
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
315 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATGCTGTGCTAGAGA 364
48 GlyAsnArgArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64
365 GGAACCGTGGAGCGGTTCCGCTTAGGCAGCCCATCTGCAGTATAGATG 414
64 pAspIleMetHisArgLeuGlyGluGluValArgGlnAlaArgMetArgGluGluA 81
415 GGACATAATGTCATAGCTTGGAGAGCCACAGCAAGGATGAGAGAGAGA 464
81 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97
465 ATATGAAAAGGATTGGGAGGAGGTGAGACAGCTGATGAAAAGCTGAGG 514
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114
```

|||||  
515 GAAAGCAGTTGAGTCATAGTCTGGGGCAGTCAGCACTGATCCCTCA 564  
|||||  
114 sHisAspHisHisAspGluPheCys.LeuMetPro 125  
|||||  
565 CCATGACCATCAGATGAGTTTGGCCCTTATGCC 599

seq\_name: /SIDS2/3cgcdata/geneseq/geneseqn/NA2000.DAT:AAC03880

seq\_documentation\_block:  
ID AAC03880 standard; cDNA; 562 BP.

XX AC AAC03880;  
XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein 5' EST, SEQ ID NO: 3878.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KW gene therapy; chromosome mapping; ss.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PA (GEST ) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI; 2000-500381/45.

XX DR P-PSDB; AAG03874.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX PS Claim 1; SEQ ID 3878; 71pp + CD-ROM; English.

XX CC The present sequence is one of a large number of 5' ESTs derived from  
XX CC mRNAs encoding secreted proteins. An ORF has been identified within the  
XX CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
XX CC derived from 30 different tissues. EST sequences usually correspond  
XX CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
XX CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
XX CC well suited for isolating cDNA sequences derived from the 5' ends of  
XX CC mRNAs and even in those cases where longer cDNA sequences have been  
XX CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
XX CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
XX CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
XX CC gene therapy and chromosome mapping procedures. They are used to obtain  
XX CC upstream regulatory sequences and to design expression and secretion  
XX CC vectors.

XX SQ Sequence 662 BP; 177 A; 144 C; 192 G; 137 T; 12 other;

alignment\_scores:  
Quality: 585.50 Length: 128  
Ratio: 4.920 Gaps: 1  
Percent Similarity: 92.969 Percent Identity: 86.719

alignment\_block:

US-09-327-750D-32 x AAC03880 ..

Align seg 1/1 to: AAC03880 from: 1 to: 662

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17

|||||  
207 ATGGAGTCCAAAGAGGAAAGGAGCGTTAAACAATCTCATCGTGGAAAATGT 256  
|||||  
17 aaasnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31  
|||||  
257 CAACCCAGGAAAATGATGAAAAAGATGAAAAAGGAGCAAGTGCCTAATAAAG 306  
|||||  
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47  
|||||  
307 GGGAGCCCTTGGCCCTTACCTTGTGATGTYRGTGAATACTGTGTGCTTAGA 356  
|||||  
48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTtr 64  
|||||  
357 GGAAATCGTAGCGGTTCCGCGTTAGGCAGCCCATCTCTGCAGTATAGATG 406  
|||||  
64 pasPileMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81  
|||||  
407 GGATATGATGATAGCTTTGGAGAACCCACAGCAAGGATGANAGAGAGA 456  
|||||  
81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97  
|||||  
457 ATATGGAAGGATTTGGGAGGAGGTGAGACAGCTGATGGAAAACCTGAGG 506  
|||||  
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114  
|||||  
507 GAAAAGCAGTTGAGTCATAGTCTGCGGGCAGTCAGCACTGACCCCTCA 556  
|||||  
114 sHisAspHisHisAspGluPheCysLeuMetPro 125  
|||||  
557 CCATGACCATCAGTATGAGTTTGCNNWATGCC 590

seq\_name: /SIDS2/3cgcdata/geneseq/geneseqn/NA2000.DAT:AAC03879

seq\_documentation\_block:

ID AAC03879 standard; cDNA; 698 BP.

XX AC AAC03879;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein 5' EST, SEQ ID NO: 3877.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KW gene therapy; chromosome mapping; ss.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PA (GEST ) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI; 2000-500381/45.

XX DR P-PSDB; AAG03873.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX PS Claim 1; SEQ ID 3877; 71pp + CD-ROM; English.

XX CC The present sequence is one of a large number of 5' ESTs derived from  
XX CC mRNAs encoding secreted proteins. An ORF has been identified within the  
XX CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
XX CC derived from 30 different tissues. EST sequences usually correspond  
XX CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.  
 XX  
 SQ Sequence 698 BP; 187 A; 151 C; 210 G; 139 T; 11 other;

alignment\_scores:  
 Quality: 585.50 Length: 128  
 Ratio: 4.920 Gaps: 1  
 Percent Similarity: 92.969 Percent Identity: 86.719

alignment\_block:

US-09-327-750D-32 x AAC03879 ..

Align seg 1/1 to: AAC03879 from: 1 to: 698

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
 |||||  
 243 ATGGAGTCCAAAGAGGAAACGAGCGTTAAACAATCTCATCGTGGAAAATGT 292  
 |||||  
 17 aasnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31  
 |||||  
 293 CAACAGGAAATATGATGAAAAGATGAAAAGGAGCAAGTGWCTTAATAAAG 342  
 |||||  
 31 lylGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47  
 |||||  
 343 GGGAGCCCTTGGCCCTACCTTGTGATGTYRGTAATCTGTGTGCCTAGA 392  
 |||||  
 48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64  
 |||||  
 393 GGAATCGTAGGCGGTTCGCGTGTAGGACGCCCATCTGCAGTATAGATG 442  
 |||||  
 64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81  
 |||||  
 443 GGATATGTCATAGTCTGCGGAGACACAGGCAAGGATCANAGAAGAGA 492  
 |||||  
 81 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97  
 |||||  
 493 ATATGGAAGGATTTGGGAGGAGCTGAGACAGCTGATGGAAGAGCTGAGG 542  
 |||||  
 98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114  
 |||||  
 543 GAAAGACAGTTGATCATAGTCTGCGGCGAGTCAGCACCTGACCCCTCA 592  
 |||||  
 114 sHisAspHisHisAspGluPheCysLeuMetPro 125  
 |||||  
 593 CCATGACCATCATGATGAGTTTGCNNWATGCC 626

seq\_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:AAC10889

seq\_documentation\_block:

ID AAC10889 standard; cDNA; 692 BP.

XX

AC AAC10889;

XX

DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 14964.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX

OS Homo sapiens.

XX

PN EP1033401-A2.

XX

PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI: 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 1; SEQ ID 14964; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX  
 SQ Sequence 692 BP; 184 A; 151 C; 199 G; 146 T; 12 other;

alignment\_scores:

Quality: 562.50 Length: 125  
 Ratio: 4.849 Gaps: 1  
 Percent Similarity: 92.800 Percent Identity: 84.800

alignment\_block:

US-09-327-750D-32 x AAC10889 ..

Align seg 1/1 to: AAC10889 from: 1 to: 692

4 LysGluLysArgAlaValAsnSerLeuSerMetGluAsnAlaAsnGlnGl 20  
 |||||  
 246 CAAAGAGAACGAGCGTTAAACAATCTCATCGTGGAAAATGTCAACCCAGGA 295  
 |||||  
 20 uasn.....GluGluLysGluGlnValAlaAsnLysGlyGluProL 34  
 |||||  
 296 AAATGATGAAAAGATGAAAAGGAGCAAGTGWCTAATAAAGGGAGCCCT 345  
 |||||  
 34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50  
 |||||  
 346 TGGCCCTACCTTGTGATGTYRGTAATCTGTCTGCTAGAGGAATCGT 395  
 |||||  
 51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrAspIleMe 67  
 |||||  
 396 AGCGCGTTCCGGCTTAGGCGAGCCCATCTGCAGTATAGATGGATATGAT 445  
 |||||  
 67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84  
 |||||  
 446 GCATAGGCTTGGAGAACCAACGAGGATGANAGAGAGATATGGANA 495  
 |||||  
 84 rgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100  
 |||||  
 496 GGATTGGGCGAGGAGGTGAGACAGCTGATGGAAGCTGAGGGAAGACAG 545  
 |||||  
 101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHl 117  
 |||||  
 546 TTGAGTCATAGTCTGCGGCGAGTCAGCACTGACCCCTCACCATTGACCA 595

117 shiAspGluPheCysLeuMetPro 125  
|||||  
596 TCATGATGAGTTTGCNNWATGCC 620

seq\_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT: AAC06183

seq\_documentation\_block:

ID AAC06183 standard; cDNA; 421 BP.

AC AAC06183;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 10258.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 10258; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

XX Sequence 421 BP; 110 A; 103 C; 124 G; 82 T; 2 other;

alignment\_scores:

Quality: 207.50 Length: 72  
Ratio: 3.402 Gaps: 2  
Percent Similarity: 84.722 Percent Identity: 65.278

alignment\_block:

US-09-327-750D-32 x AAC06183 ..

Align seg 1/1 to: AAC06183 from: 1 to: 421

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
|||||  
207 ATGGAGTCCAAAGAGGACGCGTTAAACAATCTCATCTCGTGGAAATGT 256

17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31  
:|||||  
257 CAACAGGAAATGATGAAAGATGAAGAGGACGCAAGTTGCTAATAAG 306  
|||||  
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47  
|||||  
307 GGGAGCCCTTGCCCTTACCTTTGAATGTAGTGAATAGTGTGTGCTGCTAGA 356  
|||||  
48 GlyAsnArgArgPheProValArgGlnProIleLeuGln.TyrArgT 64  
|||||  
357 GGAACCCGTAG.CGGTTCGGCTTAGGCACACAGATTATTAGAGGCCCGCT 405  
|||||  
64 rpAspIleMetHis 68  
::: |||  
406 GCCCAGTGACACAT 419

seq\_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT: AAC06180

seq\_documentation\_block:

ID AAC06180 standard; cDNA; 457 BP.

XX AAC06180;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 10255.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 10255; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

XX Sequence 457 BP; 120 A; 110 C; 142 G; 84 T; 1 other;

alignment\_scores:

Quality: 207.50 Length: 72  
Ratio: 3.402 Gaps: 2

Percent Similarity: 84.722 Percent Identity: 65.278

## alignment\_block:

US-09-327-750D-32 x AAC06180 ..

Align seg 1/1 to: AAC06180 from: 1 to: 457

```
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
243 ATGGAGTCCAAAGAGAACGCGTTAAACAATCTCATCGTGGAAATGT 292
|||||
17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
|||||
293 CAACACAGAAATGATGAAAAAGATGAAAGGACGCAAGTTGCTAATAA 342
|||||
31 lYcLupProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
|||||
343 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTCAATACTGTGTGCTAGA 392
|||||
48 GlyAsnArgArgArgPheProValArgGlnProIleLeuGln.TyrArgT 64
|||||
393 GGAACCCGTAG.CGGTTCGCGGTAGGCACACAGTATTAGAGCCACCGT 441
|||||
64 rpAspIleMethIs 68
::: |||
442 GCCCAGTGACACAT 455
```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH03517

## seq\_documentation\_block:

ID AAH03517 standard; cDNA; 865 BP.

XX AC AAH03517;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA clone (5'-primer) SEQ ID NO:352.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isodai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

XX PS Claim 1; SEQ ID 352; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX SQ Sequence 865 BP; 239 A; 177 C; 224 G; 221 T; 4 other;

## alignment\_scores:

Quality: 202.50 Length: 126  
Ratio: 2.382 Gaps: 3  
Percent Similarity: 67.460 Percent Identity: 38.889

## alignment\_block:

US-09-327-750D-32 x AAH03517 ..

Align seg 1/1 to: AAH03517 from: 1 to: 865

```
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
191 ATGGAGTCCAAAGAGAACGCGTAAACAATCTCAACGGGGAATAATCC 240
|||||
17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
241 CCAACAAGAAACGAAGGAGGAGGAGCGGCCCGCCAGCAGAGTGAAGAAG 290
|||||
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
::: |||
291 AATCCCGCCATTGGGAGGGGTGAAGGCCAGAGCCTGGAGGAATATC 340
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
|||||
341 AGCGGGGCGGAGTTAGGCGACTTGTCCCTAATTTTCGATGGCCATACC 390
|||||
67 thisArg...LeuGlyGluProGlnAlaArgMetArgGluGluAsnMetG 83
|||||
391 TAATAGGCATATTGAGCAACAATGAAGCGAGA.....GATGATGTAG 431
|||||
83 luArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLys 99
|||||
432 AAAGGTTTGTAGGGCAGATGATGGAATCAAGAGAAAGACTAGGGGAACAG 481
|||||
100 GlnLeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAs 116
|||||
482 CAGATGAGGCACATATATGCGCTTCCAAACTCTCTGAACT.....GA 522
|||||
116 pHishHisAspGluPheCysLeuMetPro 125
|||||
523 CAACCATATGACTTTTCCTCATACCT 550
```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH13750

## seq\_documentation\_block:

ID AAH13750 standard; cDNA; 1229 BP.

XX AC AAH13750;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA sequence SEQ ID NO:10656.





OS	
XX	Homo sapiens.
PN	WO200075278-A2.
XX	
PD	14-DEC-2000.
XX	
XX	07-JUN-2000; 2000WO-US15621.
PF	
XX	
PR	07-JUN-1999; 99US-0327750.
XX	
XX	
PA	(UYCO ) UNIV COLUMBIA NEW YORK.
XX	
XX	Sato T;
PI	
XX	
DR	WPI; 2001-061707/07.
XX	
XX	
PT	New p75-neurotrophin receptor-associated cell death executor (NADE) and
PT	the gene encoding NADE, useful for modulating the activity of p75NTR
PT	and for detecting neurodegenerative diseases -
XX	
PS	Disclosure; Fig 1; 134pp; English.
XX	
CC	The present invention relates to a purified polypeptide capable of
CC	binding neurotrophin receptor (p75-NTR). The invention is useful for
CC	binding and modulating the activity of p75NTR. The peptide mediates
CC	NGF-induced apoptosis, which plays an important role in neurogenetic
CC	diseases. The peptide of the invention and p75NTR are useful for
CC	inhibiting NF-kappaB activation in a cell or a subject for inducing
CC	caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase
CC	and fragment nuclear DNA in a cell by co-expression of (I) and p75-NTR.
XX	
XX	Sequence 891 BP; 251 A; 182 C; 224 G; 234 T; 0 other;
SQ	

WPI: 2001-061707/07.

**Pt** New p75-neurotrophin receptor-associated cell death executor (NADE) and  
**Pt** the gene encoding NADE, useful for modulating the activity of p75NTR  
**Pt** and for detecting neurodegenerative diseases -  
**XX**

---

**PS** Disclosure: Fig 1; 134pp; English

xx CC The present invention relates to a purified polypeptide capable of  
CC binding neurotrophin receptor (p75-NTR). The invention is useful for  
CC

CC BINDING AND MODULATING THE ACTIVITY OF p75<sup>NTR</sup> IN NGF-INDUCED APOPTOSIS, WHICH PLAYS AN IMPOR-

CC diseases. The peptide of the invention and p5NTR are useful for  
CC inhibiting NF-kappaB activation in a cell or a subject, for inducing  
CC caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase  
CC and fragment nuclear DNA in a cell by co-expression of (I) and p75-NTR.  
XX  
SQ Sequence 891 BP: 251 A; 182 C; 224 G; 234 T; 0 other:

```

alignment_scores:
  Quality: 201.00
  Ratio: 2.610
  Percent Similarity: 62.097
  Length: 124
  Gaps: 4
  Percent Identity: 43.548

alignment_block:
  US-09-327-750D-32 x AAF23529
  ..

```

```

Align seg 1/1 CO: AAF23529 From: 1 CO: 691
4 LysGluLysArg...AlaValAsnSerLeuSerMetClnAsnAlaAsnGI 19
|||||:|||||
279 ANAGAAAAACACCGAAAAAAAATCTCATCATGCGAAATATTTCACCA 328
|||||:|||||
19 nGluAsnGlnGluLysGlnValAlaAsnGlyGlyGluProLeuAlaL 36
|||||:|||||

```

36 euProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArgArgArg 52  
||||| :||| ||| |||||

378 GCGCTTTGGGAGGAGGTGAAGGCCACCGCTTGCAGGGAATCGACGGGA 423

53 PheProValArgGlnProIleLeuGlnTyrArgGTrpAspIleMetHisAr 69

426 ...CAGGCTCGCCGACTTGCCCCCTAATTTCGATGGGCCATACCCAATAG 472

69 gLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgIleG 86

473 GCAGATCAATGATGGGATGGGTGGAGATGGAGATGATATGGAAATATTTCA 522

86 102

523 TGCAGCGCATCACACAGAAATCAGACACAAACTTACCCACCTGCACCTGACG 572

103 HIGGORY CRYSTAL COVE EMBURY  
DROEDRIV CRYSTAL COVE EMBURY

[illegible]

5/3 AATTGTC TGGGTA TCCTTA TGGGGG AGCTCT CTAATC ACCATG ACCATCA 622

118 sAsnGluPheCysIleuMetPro 125

623 TGGTGAATTTGGCTTATGCCT 644

seq\_name: /SIDS2/gcgdata/geneseq/NA2000.DAT:AAF21748

seq\_documentation\_block:

ID AAF21748 standard; DNA; 917 BP.

XX AC AAF21748;

XX DT 27-MAR-2001 (first entry)

XX DE Human breast and ovarian cancer associated antigen gene SEQ ID 135.

XX KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
KW nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;  
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;  
KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
KW cardiovascular disorder; wound healing; neurological disease; ds.

XX OS Homo sapiens.

XX PN WO200055173-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05881.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX WPI; 2000-611515/58.

XX DR P-PSDB; AAB58845.

XX PT New human breast and ovarian cancer associated gene sequences and the  
PT polypeptides encoded by these genes, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -

XX PS Claim 1; Page 581-582; 1299pp; English.

XX CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
CC associated with breast and ovarian cancer. Included in the invention are  
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
CC isolation and characterisation of the DNA and protein sequences of the  
CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
CC nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
CC antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;  
CC antibacterial; antifungal; antiparasitic and cardiac activity. The  
CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
CC particularly breast and ovarian cancer. The nucleic acid sequences,  
CC proteins, agonists and antagonists may also be used in the diagnosis,  
CC prevention and treatment of immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemias; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases.

XX SQ Sequence 917 BP; 228 A; 203 C; 270 G; 211 T; 5 other;

alignment\_scores:

Quality: 185.50

Length: 108

Ratio: 2.769

Gaps: 3

Percent Similarity: 62.037 Percent Identity: 44.444

alignment\_block:

US-09-327-750D-32 x AAF21748

Align seg 1/1 to: AAF21748 from: 1 to: 917

19 GlnGluAsnGluGluLysGluGlnValAlaAsnLysGluProLeuAl 35  
|||||  
306 CAGGAAACCAAGAGATGGAGCAGCCTATGCAGAATGGAGAGAA...GA 352  
35 aLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArgArgA 52  
|||||  
353 CCGCCCTTTGGGAGGAGGTGAAGGCCACCCAGCCTGCAGGAATCGACGG 402  
52 rgPheProValArgGlnProIleLeuGlnTyrArgTTPAspIleMethis 68  
|||||  
403 GA...CAGGCTCGCCGACTTGCCTAATTTTCGATGGGCCATACCCAAT 449  
69 ArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluArgI 85  
|||||  
450 AGCCAGATCAATGATGGGATGGGTGGAGATGGAGATGATGGAATATT 499  
85 eGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGlnLeu 102  
: |||||  
500 CATGGAGGAGATGAGAGAAATCAGAGAAACTTAGGAGCTGCAGTTGA 549  
102 erHisSerLeuArgAlaValSerThrAsp...ProProHisHisAspHis 117  
: |||||  
550 GGAATTCCTCGCTATCTTATGGGAGGCTCTCTAATCACCATGACCAT 599  
118 HisAspGluPheCysLeuMetPro 125  
|||||  
600 CATGATGAATTTGCCTTATGCCT 623

**THIS PAGE BLANK (USPTO)**

OM of: US-09-327-750D-32 to: GenEmbl.\* out\_format : pfs  
Date: Mar 11, 2002 3:34 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n\_pool-US09327750/runat\_11032002\_101154\_20319/app\_query.fasta\_1.1472  
-O=/sgn2.1/USPTO\_spool/US09327750 -DEVS=xlh  
-DB=genEmbl -QWMI=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DICALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09327750 -CGNL\_1\_8673  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-32

Query length: 125

Database: GenEmbl.\*

Database sequences: 1472140

Database length: 341344837

Search time (sec): 4557.230000

score\_list:

Sequence	Strd	Orig	zscore	Escore	Len	Documentation
gb_pr:AF183416	+	653.00	1124.31	2.4e-54	642	AF183416 Homo sapiens ovarian
gb_pr:AF220189	-	653.00	1122.20	3.1e-54	828	AF220189 Homo sapiens uncharact
gb_pr:HS198P4	-	653.00	1091.01	1.7e-52	35714	AL008708 Homo sapiens DNA sequen
gb_pr:AF237783	+	639.00	1098.39	6.6e-53	791	AF237783 Homo sapiens brain-exp
gb_pr:AF251053	+	602.50	1035.82	2.0e-49	744	AF251053 Homo sapiens X-linked
gb_pat:AX078272	+	602.50	1034.27	2.5e-49	898	AX078272 Sequence 76 from Paten
gb_pr:HSV870H8	+	602.50	1004.84	1.1e-47	31321	270233 Human DNA sequence fr
gb_pr:AL133348	-	602.50	1002.69	1.4e-47	40584	AL133348 Human DNA sequence fr
gb_htg:HSU08081	+	602.50	1002.60	1.4e-47	41029	AL022169 Homo sapiens chromos
gb_pr:AF097439	+	496.00	851.36	3.8e-39	785	AF097439 Mus musculus brain exp
gb_pr:AF097437	+	448.50	760.49	4.4e-34	2269	AF097437 Mus musculus Bex1 pr
gb_pr:AF097438	+	446.50	765.31	2.4e-34	835	AF097438 Mus musculus brain exp
gb_pr:AF051347	+	446.50	765.28	2.4e-34	838	AF051347 Mus musculus brain exp
gb_pr:BC003254	+	384.50	647.86	8.3e-28	2901	BC003254 Mus musculus, Similar
gb_sts:G24641	-	348.00	599.29	4.2e-25	504	G24641 human STS WI-11354, sequ
gb_pr:HSV351F8	-	216.50	334.73	2.3e-10	45678	270719 Human DNA sequence fr
gb_pr:AK000959	+	202.50	340.49	1.1e-10	1229	AK000959 Homo sapiens cDNA FL
gb_pat:AX100231	+	202.50	339.63	1.2e-10	1364	AX100231 Sequence 32 from Pat
gb_pr:HS635G19	+	202.50	307.05	7.9e-09	69648	AL035494 Human DNA sequence f
gb_pr:HUMOGC	+	201.00	340.57	1.1e-10	891	M38188 Human unknown protein fr
gb_pr:AF187064	+	201.00	340.57	1.1e-10	891	AF187064 Homo sapiens p75NTR-as
gb_pr:BC003190	+	199.50	338.94	1.3e-10	793	BC003190 Homo sapiens, p75NTR-a
gb_pr:AF187066	+	186.00	316.65	2.3e-09	700	AF187066 Mus musculus p75NTR-as
gb_pr:AF097440	+	186.00	315.00	2.9e-09	854	AF097440 Mus musculus brain exp
gb_sts:G35294	-	170.00	292.18	5.4e-08	477	G35294 human STS SHGC-37409, se
gb_pr:AF187065	+	169.00	289.75	7.3e-08	519	AF187065 Rattus norvegicus p75N
gb_sts:G27208	-	138.50	238.78	5.0e-05	421	G27208 MARC 4953-4954:991939031
gb_sts:G23964	-	103.50	179.33	0.1033	372	G23964 human STS WI-15922, sequ
gb_om:AF023169	+	97.50	147.57	6.07	4923	AF023169 Canis familiaris type
gb_om:MMU05823	+	89.00	130.03	57.60	6945	U05823 Mus musculus pericentri
gb_pr:HSU31906	+	89.00	130.00	57.77	6965	X13906 Homo sapiens golgin-245
gb_pr:HSUGOLGIN	+	89.00	129.19	64.12	7693	X28234 H.sapiens mRNA for golg
gb_pr:HSU41740	+	89.00	129.18	64.23	7695	U41740 Human trans-Golgi p230
gb_ba:AE000565	+	89.00	125.65	100.94	11776	AE000565 Helicobacter pylori
gb_pr:AL136998	+	88.00	100.96	2.4e+03	186218	AL136998 Mouse DNA sequence
gb_pr:AB062945	+	87.00	139.98	16.06	1376	AB062945 Macaca fascicularis b
gb_pl:AF332874	+	86.50	135.20	29.65	2207	AF332874 Oryza sativa phosphol
gb_pr:RATNESTIN	+	86.50	126.99	84.99	5946	M34384 Rat nestin mRNA, comple
gb_om:AF332860S2	+	86.50	104.48	1.5e+03	89994	AF332861 Mus musculus diadend
gb_ba:LDMOFA	-	86.00	122.32	154.69	9413	225774 Leptothrix discophora m

gb\_htg:AC068585 + 86.00 105.66 1.3e+03 70358 AC068585 Homo sapiens clon  
gb\_htg:AC093313 - 86.00 100.00 2.7e+03 139315 AC093313 Trypanosoma cruz  
gb\_ro:AB041591 - 85.00 134.99 30.49 1657 AB041591 Mus musculus brain  
gb\_htg:AC013960 + 85.00 102.53 2.0e+03 83321 AC013960 Drosophila melano  
gb\_in:AC009346 + 85.00 96.70 4.1e+03 169448 AC009346 Drosophila melar

seq\_name: gb\_pr:AF183416

seq\_documentation\_block:

LOCUS AF183416 642 bp mRNA PRI 02-SEP-2000  
DEFINITION Homo sapiens ovarian granulosa cell 13.0 kDa protein hGR74 homolog  
mRNA, complete cds.

ACCESSION AF183416

VERSION AF183416.1 GI:9963770

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 642)

AUTHORS Peng, Y., Qian, B., Tu, Y., Xu, S., Han, Z., Fu, G. and Chen, Z.

TITLE A novel gene expressed in human adrenal gland

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 642)

AUTHORS Peng, Y., Gu, W., Huang, C., Xu, S., Han, Z., Fu, G. and Chen, Z.

TITLE Direct Submission

JOURNAL Submitted (03-SEP-1999) Chinese National Human Genome Center at  
Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,  
Shanghai 201203, P. R. China

FEATURES

source

1..642

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/tissue\_type="adrenal gland"

3..380

/note="HGR74-h protein"

/codon\_start=1

/evidence=not\_experimental

/product="ovarian granulosa cell 13.0 kDa protein hGR74

homolog"

/protein\_id="AAG09685.1"

/db\_xref="GI:9963771"

/translation="MESKEKRAVSLSMENAOENEEKEQVANKGEPLALPLDAGEYCV  
VPRGNRRFRVROPILQYRWDMHRLGEPQARMRENMERIGEVRQLMEKLRLKRLS  
HSLRAVSTDPHPHDHDECLMP"

BASE COUNT 204 a 118 c 157 g 163 t

ORIGIN

alignment\_scores:

Quality: 653.00 Length: 125

Ratio: 5.266 Gaps: 0

Percent Similarity: 99.200 Percent Identity: 98.400

alignment\_block:

US-09-327-750D-32 x AF183416

Align seg 1/1 to: AF183416 from: 1 to: 642

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17

|||||

3 ATGCAGTCCAAAGAGAAACGAGCAGTAAACAGCTCAGCAGTGAATGC 52

|||||

17 aAsnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34

|||||

53 CAACCAAGAAAATGAAGAAAGGAGCAGGTTGCTTAATAAAGGGGGCCCT 102

|||||

34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50

|||||

103 TGGCCCCCCTTTGGTGGTGGTGAATCTGTGTGCTTAGAGGAATCGT 152

|||||

51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67

|||||

```

153 AGGCGGTTCCGGTTAGCAGCCCATCTCGCAGTATAGATGGGATATGAT 202
67  tHisArgLeuGluGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
   |||||||
203 GCATAGGCTTCGAGNACCACAGGCAAGGATGAGAGAGAAATATGGNAA 252
   |||||||
84  rGileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
   |||||||
253 GGATTGGGAGGAGGTGAGACAGCTGATGGAAAGCTGAGGGAAGACAG 302
   |||||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117
   |||||||
303 TTGAGTCATAGTCTCGCGGCGAGTCAGCACTGACCCCTCACCATGACCA 352
   |||||||
117 sHisAspGluPheCysLeuMetPro 125
   |||||||
353 TCATGATGAGTTTTCCTTATGCCC 377

```

seq\_name: gb\_pr:AF220189

```

seq_documentation_block:
LOCUS AF220189 828 bp mRNA PRI 04-MAY-2000
DEFINITION Homo sapiens uncharacterized hypothalamus protein HBEX2 mRNA,
complete cds.
ACCESSION AF220189
VERSION AF220189.1 GI:7689028
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.
JOURNAL A novel gene expressed in human hypothalamus
Unpublished
REFERENCE 2 (bases 1 to 828)
AUTHORS Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Chinese National Human Genome Center at
Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, P. R. China
FEATURES
source
1..828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="hypothalamus"
167..544
/codon_start=1
/product="uncharacterized hypothalamus protein HBEX2"
/protein_id="AAF67654.1"
/db_xref="GI:7689029"
/translation="WESKEKRAVNSLSMENAOENEEKVOVANKGEPLALPLDAGEYCV
VRGNRRFRFRVQPIQYRWDMHRLGEPQARMRENERMERIGEEVRLMEKLRKQLS
HSLRAVSTDPDPPHHDDHDFCLMP"
BASE COUNT 256 a 170 c 221 g 181 t
ORIGIN

```

alignment\_scores:  
Quality: 653.00 Length: 125  
Ratio: 5.266 Gaps: 0  
Percent Similarity: 99.200 Percent Identity: 98.400

alignment\_block: \*  
US-09-327-750D-32 x AF220189 ..

Align seg 1/1 to: AF220189 from: 1 to: 828

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
   |||||||
167 ATGAGTCCCAAGAAACGACGACGTAACAGCTCAGCATGGGAAATGC 216
   |||||||
17 aAsnGlnGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 34

```

```

217 CAACCAAGAAAATGAGAAAAGGAGCAAGTTGCTAATAAAGGGAGCCCT 266
   |||||||
34 euAlaLeuProLeuAspAlaGlyClyTyrcysValProArgGlyAsnArg 50
   |||||||
267 TGGCCCTCCCTTTGGATGCTGGTGAATACTGTGTGCCTAGAGGAAATCGT 316
   |||||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyArgTrpAspIleMe 67
   |||||||
317 AGCGGTTCGCGTTAGCGACCCATCTCGCAGTATAGATGGGATATGAT 366
   |||||||
67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
   |||||||
367 GCATAGGCTTGGAGAACACACAGGCAAGGATGAGAGAGAGAAATATGAAA 416
   |||||||
84 rGileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
   |||||||
417 GGATTGGGAGGAGGTGAGACAGCTGATGGAAAGCTGAGGGAAGACAG 466
   |||||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117
   |||||||
467 TTGAGTCATAGTCTCGCGGCGAGTCAGCACTGACCCCTCACCATGACCA 516
   |||||||
117 sHisAspGluPheCysLeuMetPro 125
   |||||||
517 TCATGATGAGTTTTCCTTATGCCC 541

```

seq\_name: gb\_pr:HS198P4

```

seq_documentation_block:
LOCUS HS198P4 35714 bp DNA PRI 16-DEC-2000
DEFINITION Human DNA sequence from clone RPI-198P4 on chromosome Xq22 Contains
a CpG Island, complete sequence.
ACCESSION AL008708
VERSION AL008708.4 GI:4826445
KEYWORDS HTG; CpG Island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Corby,N.
JOURNAL Direct Submission
Submitted (21-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 13, 1999 this sequence version replaced gi:4582117.
COMMENT
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
RPI-198P4 is
from the library RPCI-1 constructed at the Roswell Park Cancer
Institute by the group of Pieter de Jong. For further details see
http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone

```

Percent Similarity: 91.667 Percent Identity: 84.722

# alignment\_block:

US-09-327-750D-31 x AAC06180

Align seg 1/1 to: AAC06180 from: 1 to: 457

```

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
243 ATGGAGTCCAAAGAGAGAGCGGTTAAACATCTCATCGTGGAAATGT 292
|||||
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
293 CAACACAGGAAATGATGAAAAAGATGAAAAAGGAGCAAGTTGCTAATAAAG 342
|||||
34 lYcGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
343 GGGAGGCCCTGGGCCCTTACCTTTGAATGTTAGTGAATACTGTGTGCTAGA 392
|||||
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGln.TyrArgT 67
|||||
393 GGAACCCGTAG.CGGTTCCGCGTTAGGCACACAGTATTAGAGCCACCGCT 441
|||||
67 rpAspIleMethis 71
::: |||
442 GCCCAGTGACACAT 455

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAC06184

# seq\_documentation\_block:

ID AAC06184 standard; cDNA; 451 BP.

AC AAC06184;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 10259.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 10259; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively  
XX identified within the present sequence. The 5' ESTs were prepared from  
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
XX sequences usually correspond mainly to the 3' untranslated region (UTR)  
XX of the mRNA because they are often obtained from oligo-dT primed cDNA  
XX libraries. Such ESTs are not well suited for isolating cDNA sequences  
XX derived from the 5' ends of mRNAs and even in those cases where longer  
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.  
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

SQ Sequence 451 BP; 117 A; 110 C; 131 G; 91 T; 2 other;

# alignment\_scores:

Quality: 261.00 Length: 69  
Ratio: 4.143 Gaps: 1  
Percent Similarity: 91.304 Percent Identity: 81.159

# alignment\_block:

US-09-327-750D-31 x AAC06184

Align seg 1/1 to: AAC06184 from: 1 to: 451

```

4 LysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnValAsnGlnG 20
|||||
246 CAAGAGAGACGAGCGTTAAACATCTCATCGTGGAAATGTCAACACAGA 295
|||||
20 uAsnAspGluLysAspGluLysGluGlnValAlaAsnLysGluProL 37
|||||
296 AAATGATGAAAAAGATGAAAAAGGAGCAAGTTGCTAATAAAGGGAGCCT 345
|||||
37 euAlaLeuProLeuAsnValSerGluTyrCysValProArgGlyAsnArg 53
|||||
346 TGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCTAGAGGAACCGT 395
|||||
54 ArgArgPheArgValArgGlnProIleLeuGln.TyrArgTyrAspIle 70
|||||
396 AG.CGGTTCCGCGTTAGGCACACAGTATTAGAGGCACCGCTGCCAGTGA 444
|||||
70 ethis 71
|||
445 CACAT 449

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAC23529

# seq\_documentation\_block:

ID AAC23529 standard; DNA; 891 BP.

XX AAC23529;

XX 22-MAR-2001 (first entry)

XX Human NADE DNA.

XX Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;  
KW neurogenetic disease; NF-kappaB; ds.

XX Homo sapiens.

XX WO200075278-A2.

XX 14-DEC-2000.

XX 07-JUN-2000; 2000WO-US15621.

XX 07-JUN-1999; 99US-0327750.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

XX Sato T;

XX WPI; 2001-061707/07.

XX New p75-neurotrophin receptor-associated cell death executor (NADE) and  
XX the gene encoding NADE, useful for modulating the activity of p75NTR  
XX and for detecting neurodegenerative diseases -

XX Disclosure; Fig 1; 134pp; English.

117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
 |||||  
 514 CCATGACCATCATGATGAGTTTGCCTTATGCC 547

seq\_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:AAC06183

seq\_documentation\_block:

ID AAC06183 standard; cDNA: 421 BP.

XX AAC06183;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 10258.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPT; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 10258; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

XX Sequence 421 BP; 110 A; 103 C; 124 G; 82 T; 2 other;

alignment\_scores:

Quality: 284.00 Length: 72  
 Ratio: 4.303 Gaps: 1  
 Percent Similarity: 91.667 Percent Identity: 84.722

alignment\_block:

US-09-327-750D-31 x AAC06183 ..

Align seg 1/1 to: AAC06183 from: 1 to: 421

1 MetGluSerLysGluGluArgAlaLeuAsnLeuValGluAsnVa 17  
 |||||  
 207 ATGGAGTCCAAAGAGGACGCGTTAAACAATCTCATCGTGGAAATGCT 256

17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34  
 |||||  
 257 CAACACAGGAATGATGAAAAAGATGAAAAAGGAGCAAGTTGCTAATAAG 306  
 |||||  
 34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
 |||||  
 307 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCTAGA 356  
 |||||  
 51 GlyAsnArgArgPheArgValArgGlnProIleLeuGln.TyrArgT 67  
 |||||  
 357 GGAACCCGTAG.CGGTTCCGCGTTAGGCACACAGATTATTAGAGCACCGCT 405  
 |||||  
 67 rpAspIleMethis 71  
 :::: ||||  
 406 GCCCAGTGACACAT 419

seq\_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:AAC06180

seq\_documentation\_block:

ID AAC06180 standard; cDNA: 457 BP.

XX AAC06180;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 10255.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPT; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 10255; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

XX Sequence 457 BP; 120 A; 110 C; 142 G; 84 T; 1 other;

alignment\_scores:

Quality: 284.00 Length: 72  
 Ratio: 4.303 Gaps: 1



CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

XX  
 SQ Sequence 692 BP; 184 A; 151 C; 199 G; 146 T; 12 other;

alignment\_scores:  
 Quality: 626.00 Length: 125  
 Ratio: 5.174 Gaps: 0  
 Percent Similarity: 96.800 Percent Identity: 93.600

alignment\_block:  
 US-09-327-750D-31 x AAC10889 ..

Align seg 1/1 to: AAC10889 from: 1 to: 692

4 LysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnValAsnGlnG1 20

246 CAAAGAGACGCGTAAACATCTCATCGTGGAAATGTCACACGAGA 295

20 uAsnAspGluLysAspGluGlnValAlaAsnLysGlyGluProL 37

296 AATGATGAAAAGATGAAAGGAGCAGTGWCTATAAGGGGAGCCCT 345

37 euAlaLeuProLeuAsnValSerGluTyrCysValProArgGlyAsnArg 53

346 TGGCCCTACCTTTGRTATGTGTGAATCTGTGTGCTAGAGGAAATCGT 395

54 ArgArgPheArgValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 70

396 AGCGGTTCCGCTTAGGAGCCCTCTGCAGTATAGATGATGATATGAT 445

70 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 87

446 GCATAGCTTGGAGAACACAGCAGGAGGATGANAAGAGAGAAATATGAAA 495

87 rgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 103

496 GGATGGGGAGGAGGTGAGACAGCTGATGAAAAGCTGAGGGAAGACAG 545

104 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 120

546 TTGAGTCATAGTCTGGGGAGTCACACACTGACCCCTCACCCTCACCACCA 595

120 sHisAspGluPheCysLeuMetPro 128

596 TCATGATGAGTTTTCGNNWATGCC 620

seq\_name: /SID62/gcgdata/geneseq/geneseq/NA2001.DAT:AAH45143

seq\_documentation\_block:

ID AAH45143 standard; cDNA: 792 BP.

XX AC AAH45143;

XX DT 07-SEP-2001 (first entry)

XX DE Human brain expressed x-linked protein, hBex, coding sequence.

XX KW Human; brain expressed x-linked protein; cytostatic; auditory; nontropic;

XX KW hBex; dysembryoplasia; hereditary disease; cancer; tumour; deafness;

XX KW x-chromosome-binding mental retardation; lissencephalous disease; ss.

XX Homo sapiens.  
 OS WO200140286-A1.  
 XX  
 PN 07-JUN-2001.  
 XX  
 PD  
 XX  
 PF 27-NOV-2000; 2000WO-CN00502.  
 XX  
 PR 30-NOV-1999; 95CN-0124179.  
 XX  
 PA (BIOR-) BIORAD GENE DEV LTD SHANGHAI.

XX PI Mao Y, Xie Y;

XX WIPI: 2001-397944/42.

DR P-PSDB; AAB99224.

XX Isolated human brain-expressed X-linked polypeptide used to diagnose  
 PT and treat of dysembryoplasia, hereditary diseases, cancer, tumor,  
 PT deafness and x-chromosome-binding mental retardation

XX PS Claim 5; Page 22; 30pp; Chinese.

XX The present sequence is the coding sequence for a human brain-expressed  
 CC x-linked protein (hBex). hBex and its coding sequence are useful in the  
 CC diagnosis and treatment of dysembryoplasia, hereditary diseases, cancer,  
 CC tumours, deafness, x-chromosome-binding mental retardation and  
 CC lissencephalous disease. hBex is also useful for screening mimics,  
 CC agonists, or inhibitors, and in peptide fingerprinting identification.  
 CC hBex coding sequence can be used as primers or probes, or in producing  
 CC gene chips or microarrays.

XX SQ Sequence 792 BP; 214 A; 172 C; 219 G; 187 T; 0 other;

alignment\_scores:  
 Quality: 592.50 Length: 128  
 Ratio: 4.857 Gaps: 1  
 Percent Similarity: 95.312 Percent Identity: 87.500

alignment\_block:

US-09-327-750D-31 x AAH45143 ..

Align seg 1/1 to: AAH45143 from: 1 to: 792

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17

173 ATGGAGTCCAAAGAGAAACTAGCAGTAAACAGTCTCAGCATGGAAAATGC 222

17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34

223 CAACCAAGAAAAAT.....GAAGAAAAGGAGCAAGTTGCTAATAAAG 263

34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50

264 GGGAGCCCTTGGCCCTCCCTTTGGATGCTGCTGAATACTGTGTGCCTAGA 313

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67

314 GGAATCGTAGGCGGTTCGCCGTTAGGACGCCCATCTGCAGTATAGATG 363

67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84

364 GGATATGATGATAGGCTTGGAGAACCCACAGGCAAGGATGAGAGAGAGA 413

84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100

414 ATATGGAAGGATTGGGAGGGGTGAGACAGCTGATGGAAAAGCTGAGG 463

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117

464 GAAAAGCAGTTGATGATCATAGTCTCGCGGCGAGTCAGCACTGACCCCTCA 513

84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100  
 |||||  
 457 ATATGGAAGCATGGGAGAGGTGAGACAGCTGATGGAAGAGCTGAGG 506  
 |||||  
 101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117  
 |||||  
 507 GAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCAGCTGACCCCTCA 556  
 |||||  
 117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
 |||||  
 557 CCATGACCATCATGATGAGTTTTCNNWATGCC 590  
 |||||

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC03879

seq\_documentation\_block:  
 ID AAC03879 standard; cDNA; 698 BP.

XX AAC03879;  
 AC  
 DT  
 XX 06-OCT-2000 (first entry)  
 XX

DE Human secreted protein 5' EST, SEQ ID NO: 3877.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX Homo sapiens.

OS

XX EP1033401-A2.

PN

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR P-PSDB; AAG03873.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 3877; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.

XX Sequence 698 BP; 187 A; 151 C; 210 G; 139 T; 11 other;

alignment\_scores:  
 Quality: 649.00 Length: 128  
 Ratio: 5.234 Gaps: 0  
 Percent Similarity: 96.875 Percent Identity: 95.312

alignment\_block:

US-09-327-750D-31 x AAC03879  
 Align seg 1/1 to: AAC03879 from: 1 to: 698  
 1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17  
 |||||  
 243 ATGAGGTCCAAAGAGGAAAGCGGTTAAACAATCTCATCGTGGAAATGT 292  
 |||||  
 17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34  
 |||||  
 293 CAACCCAGGAAATCATCAAAAGATGAAAAGAGGACCAAGTGCCTAATAAAG 342  
 |||||  
 34 lYGlupProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
 |||||  
 343 GGGAGCCCTTGGCCCTTACCTTGTGATGTYTRGTGAATACTGTGTCCTAGA 392  
 |||||  
 51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67  
 |||||  
 393 GGAATCGTAGCGGTTCCGCGTTAGCAGCCCATCTGCAGTATAGATG 442  
 |||||  
 67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84  
 |||||  
 443 GGATATGATGATAGCTTGGAGAACACACAGGCAAGGATGANAGAAGAGA 492  
 |||||  
 84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100  
 |||||  
 493 ATATGGAAGGATTGGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 542  
 |||||  
 101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117  
 |||||  
 543 GAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCAGCTGACCCCTCA 592  
 |||||  
 117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
 |||||  
 593 CCATGACCATCATGATGAGTTTTCNNWATGCC 626  
 |||||

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC10889

seq\_documentation\_block:

ID AAC10889 standard; cDNA; 692 BP.

XX AAC10889;  
 AC

XX 06-OCT-2000 (first entry)  
 XX

XX Human secreted protein 5' EST, SEQ ID NO: 14964.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX Claim 1; SEQ ID 14964; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX

SQ Sequence 858 BP; 236 A; 175 C; 241 G; 206 T; 0 other;

#### alignment\_scores:

Quality: 666.00 Length: 129  
 Ratio: 5.203 Gaps: 1  
 Percent Similarity: 99.225 Percent Identity: 99.225

#### alignment\_block:

US-09-327-750D-31 x AAI60367 ..

Align seg 1/1 to: AAI60367 from: 1 to: 858

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuValGluAsnVa 17  
 215 ATGGAGTCCAAAGAGAAAGAGCGGTTAAACAATCTCATCTGCGGAAATGT 264  
 17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34  
 265 CAACACAGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 314  
 34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
 315 GGCAGCCCTGGGCCCTACCTTTGAATGTTAGTGAATACTGTGTCCTAGA 364  
 51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67  
 365 GGAACCGTAGGCGGTTCGCGCTAGGACGCCATCTGCAGTATAGATG 414  
 67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 84  
 415 GGACATAATGATAGCTTTGGAGAGCCACAGGCAAGGATGAGAGAGGAGA 464  
 84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100  
 465 ATATGAAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGG 514  
 101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117  
 515 GAAAGACAGTTGAGTCATATAGTCTCGCGGCGAGTCAGCACTGATCCCCCTCA 564  
 117 sHisAspHisHisAspGluPheCys LeuMetPro 128  
 565 CCATGACATCAGATGAGTTTGGCCCTTATGCCC 599

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT: AAC03880

#### seq\_documentation\_block:

ID AAC03880 standard; cDNA; 662 BP.  
 XX  
 AC AAC03880;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 3878.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

gene therapy; chromosome mapping; ss.

Homo sapiens.

PN EPI033401-A2.

XX 06-SEP-2000.

PD 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

PR (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR P-PSDB; AAG03874.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX Claim 1; SEQ ID 3878; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.

SQ Sequence 662 BP; 177 A; 144 C; 192 G; 137 T; 12 other;

#### alignment\_scores:

Quality: 649.00 Length: 128  
 Ratio: 5.234 Gaps: 0  
 Percent Similarity: 96.875 Percent Identity: 95.312

#### alignment\_block:

US-09-327-750D-31 x AAC03880 ..

Align seg 1/1 to: AAC03880 from: 1 to: 662

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuValGluAsnVa 17  
 207 ATGGAGTCCAAAGAGAAAGAGCGGTTAAACAATCTCATCTGCGGAAATGT 256  
 17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34  
 257 CAACACAGAAATGATGAAAAAGATGAAAGGAGCAAGTGCCTAATAAAG 306  
 34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
 307 GGCAGCCCTGGGCCCTACCTTTGATGCTGCTGAATACTGTGTCCTAGA 356  
 51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67  
 357 GGAATCGTAGGCGGTTCCGCGTTAGGACGCCCATCTGCAGTATAGATG 406  
 67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 84  
 407 GGATATGATGATAGCTTTGGAGAACCCACAGGCAAGGATGANAGAAGAGA 456

KW menstrual cycle disorder; bacterial infection; ss.  
XX Homo sapiens.  
XX WO200107471-A2.  
XX 01-FEB-2001.  
XX 21-JUL-2000; 2000WO-US19948.  
XX 21-JUL-1999; 99US-0145075.  
XX 08-SEP-1999; 99US-0153129.  
XX 10-NOV-1999; 99US-0164647.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;  
PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;  
XX WPI: 2001-112727/12.  
XX P-PSDB: AAB60474.  
XX Human cell cycle and proliferation proteins and polynucleotides are  
PT used to treat, diagnose and prevent immune, developmental and cell  
PT signaling disorders and cell proliferative disorders including cancer -  
XX Claim 5; Page 181-182; 205pp; English.  
XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human  
CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.  
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
CC associated with decreased expression of functional CCYPR, while CCYPR  
CC antagonists are used to treat diseases or conditions associated with  
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
CC that specifically bind to CCYPR, and in drug screening methods to  
CC identify compounds that modulate the activity of CCYPR. CCYPR  
CC nucleotides can be used to generate transgenic animal models of human  
CC disease, and can be used in gene therapy in target cells with genetic  
CC abnormalities with respect to the expression of CCYPR for the  
CC treatment or prevention of a disorder associated with CCYPR.  
CC Diseases which can be diagnosed, treated and prevented using CCYPR  
CC proteins, nucleic acids, agonists or antagonists include immune,  
CC developmental and cell signalling disorders, and cell proliferative  
CC disorders including cancer. Specific examples of these disorders  
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
CC diabetes mellitus, disorders of the menstrual cycle and infections  
XX caused by bacteria.  
XX Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;

alignment\_scores:  
Quality: 679.00 Length: 128  
Ratio: 5.305 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750d-31 x AAF59611 ..  
Align seg 1/1 to: AAF59611 from: 1 to: 898  
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuValGluAsnVa 17  
254 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCGTGGAATAATGT 303  
17 lAsnGlnGluAsnAspGluLysAspGluLysGluValAlaAsnLysG 34  
304 CAACGAGGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTATAAAG 353  
34 lYgluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50

|||||  
354 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACCTGTGTGCCTAGA 403  
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67  
|||||  
404 GGAACCGTAGCGGTTCCGCGTTAGGCAGCCCATCTGCAGCTATAGATG 453  
67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGluA 84  
|||||  
454 GGACATAATGCATAGGCTTGGAGAGCCACAGGCAAGGATGAGAGGAGA 503  
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100  
|||||  
504 ATATGAAAGGATTGGGAGGAGGTGAGACACCTGATGAAAGCTGAGG 553  
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117  
|||||  
554 GAAAGCAGTTGAGTCATAGTCTGCGGCGAGTCAGCACCTGATCCCCCTCA 603  
117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||  
604 CCATGACCATCACGATGAGTTTGGCTTATGCC 637

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI60367

seq\_documentation\_block:

ID AAI60367 standard; cDNA; 858 BP.

XX AC AAI60367;

XX DT 22-Oct-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 4356.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-prager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI: 2001-442253/47.

XX DR P-PSDB: AAM41211.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -  
XX Claim 1; SEQ ID NO 4356; 10078pp; English.  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

251 GGAACCGTAGCGGTTCCGGTTAGCAGCCATCTCGCAGTATAGATG 300  
67 pAspIleMetHisArgLeuGluProGlnAlaArgMetArgGluAla 84  
|||||  
301 GGACATATGATGAGTGGAGACCCACAGGCAAGGATGAGAGAGAGA 350  
84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100  
|||||  
351 ATATGGAAGGATTGGGAGAGGTGACAGCTGATGGAAGAGCTGAGG 400  
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProHi 117  
|||||  
401 GAAAGCAGTTGATGATCTCTCGGGCAGTCAGCAGCTGATCCCCCTCA 450  
117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||  
451 CCATGACCATCAGCAGTGGTTTGCCTTATGCC 484

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI58581

seq\_documentation\_block:  
ID AAI58581 standard; cdna; 862 BP.

XX AAI58581;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 784.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM39425.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

XX Claim 1; SEQ ID NO 784; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX SQ Sequence 862 BP; 235 A; 183 C; 238 G; 206 T; 0 other;

alignment\_scores:  
Quality: 679.00 Length: 128  
Ratio: 5.305 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-31 x AAI58581

Align seg 1/1 to: AAI58581 from: 1 to: 862

1 MetGluSerLysGluArgAlaLeuAsnLeuIleValGluAsnVa 17  
|||||  
226 ATGGAGTCCAAAGAGGAACGAGCGCTAAACAATCTCATCTGCGAAATGT 275  
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34  
|||||  
276 CAACGAGGAAATGATGAAAAAGATGAAAAGGAGCAAGTTGCTAATAAAG 325  
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
|||||  
326 GGGAGCCCTTGGCCCTACCTTTGAATGTAGTGAATACTGTGTCCTAGA 375  
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67  
|||||  
376 GGAACCGTAGCGGTTCCGGTTAGCAGCCCATCTCGCAGTATAGATG 425  
67 pAspIleMetHisArgLeuGluProGlnAlaArgMetArgGluAla 84  
|||||  
426 GGACATATGATGAGTGGAGACCCACAGGCAAGGATGAGAGAGAGA 475  
84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100  
|||||  
476 ATATGGAAGGATTGGGAGAGGTGAGACAGCTGATGGAAGAGCTGAGG 525  
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProHi 117  
|||||  
526 GAAAGCAGTTGATGATCTCTCGGGCAGTCAGCAGCTGATCCCCCTCA 575  
117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||  
576 CCATGACCATCAGCAGTGGTTTGCCTTATGCC 609

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF59611

seq\_documentation\_block:

ID AAF59611 standard; cdna; 898 BP.

XX AAF59611;

XX 24-APR-2001 (first entry)

XX Human cell cycle and proliferation protein CCYPR-22 cdna, SEQ ID NO:76.

XX Cell cycle and proliferation protein; CCYPR; human; agonist;  
KW antagonist; gene therapy; detection; gene therapy;  
KW transgenic animal disease model; immune disorder;  
KW developmental disorder; cell signalling disorder;  
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
KW arteriosclerosis; asthma; allergy; diabetes mellitus;





SEQUENCE CHARACTERISTICS:  
LENGTH: 2576 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9..2564  
OTHER INFORMATION: /note= "Maize optimized sequence  
encoding VIPIA(a) with the Bacillus secretion signal removed  
contained in pCIB5526"  
US-08-469-334-35

alignment\_scores:  
Quality: 87.00 Length: 126  
Ratio: 1.338 Gaps: 6  
Percent Similarity: 51.587 Percent Identity: 27.778

alignment\_block:

US-09-327-750D-31 x US-08-469-334-35 ..

Align seg 1/1 to: US-08-469-334-35 from: 1 to: 2576

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31  
833 GAACGTGAGCATGGAGAGGTGATCTGTAGCCCCAACGAGACCTGAGCA 882  
31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyCysV 48  
883 ACAGCGTGGAGACCCACTCGAGCACCACACTGGAGCTACACCAACAC... 928  
48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64  
929 .....CGAGGCCCGCCAGCGTGGAGCGCGCATCGGTCCCAAGGGCATCAG 973  
65 TyrArgTrpAsp..... 68  
974 CTTCCGCGTGAGCGTGAACCTACCAGCAGCAGCGACCGTGGCCCGAGGAGT 1023  
69 .....IleMetHisArgLeuGlyGluPro...GlnAlaArgMetArgGluG 83  
1024 GGGGCACCGACCGCGGACACACCGCGCGGTTCACACCGCGCGCGC... 1069  
83 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeu 99  
1070 .....CGGCTACCTGACCGCCACCGTGGCTGCTACACACACGTGGG 1108  
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSe 112  
1109 CACCGCGCGCATCTAGCAGCTGAAGCCACCGACCGACCGCTTCGTGCT.... 1153  
112 rThrAspProProHisHisHisHisHis 121  
1154 ....GAACAACGACACCATCGCCACCAT 1177

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-300-529-35

seq\_documentation\_block:  
Sequence 35, Application US/09300529  
Patent No. 6066783  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B

APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: Genes Encoding Insecticidal Proteins  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 6066783artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/300,529  
FILING DATE: TBA  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/469,334  
FILING DATE: 06-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Meligs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-19506L  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2576 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9..2564  
OTHER INFORMATION: /note= "Maize optimized sequence  
encoding VIPIA(a) with the Bacillus secretion signal removed  
contained in pCIB5526"  
US-09-300-529-35

alignment\_scores:  
Quality: 87.00 Length: 126  
Ratio: 1.338 Gaps: 6  
Percent Similarity: 51.587 Percent Identity: 27.778

alignment\_block:

US-09-327-750D-31 x US-09-300-529-35 ..

Align seg 1/1 to: US-09-300-529-35 from: 1 to: 2576

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31  
833 GAACGTGAGCATGGAGAGGTGATCTGTAGCCCCAACGAGACCTGAGCA 882











```
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSe 112
:.....:|||||
:.....:|||||
546 CACCGCGGCATCTACGAGTGAAGCCACCCAGCTTCGTGCT..... 590
112 rThrAspProHisHisAspHis 121
:..:|||||
591 ....GAACAACGACACCATCGCCACCAT 614
```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-300-529-18

```
seq_documentation_block:
; Sequence 18, Application US/09300529
; Patent No. 6066783
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6066783artis Corporation
; STREET: 3054 Cordwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/300,529
; FILING DATE: TBA
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/469,334
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: S-19506L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
; US-09-300-529-18
```

```
alignment_scores:
  Quality: 87.00      Length: 126
  Ratio: 1.338      Gaps: 6
  Percent Similarity: 51.587      Percent Identity: 27.778

alignment_block:
US-09-327-750D-31 x US-09-300-529-18 ..

Align seg 1/1 to: US-09-300-529-18 from: 1 to: 2004

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGlnValAl 31
|||: : : |||: : : : : : : :
270 GAACGTGACATGGAGAAGGTGATCTGTAGCCGCCAACGAGACCTGAGCA 319
: : : : : : : : : : : : : :
31 aasnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysV 48
: : : : : : : : : : : : : :
320 ACAGCGTGGAGAGCCACTCGAGCACCAACTGGAGCTACACCAACAC... 365
: : : : : : : : : : : : : :
48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64
|||: : : |||: : : : : : :
366 .....CGAGGGCCGACGGTGGAGCGCGGATCGTCCCAAGGGCATCAG 410
|||: : : |||: : : : : : :
65 TyrArgTrpAsp..... 68
|||: : :
411 CTTCGCGGTGAGCGTGAACCTACCAGCACGACGCGTGGCCAGGAGT 460
69 .....IleMetHisArgLeuGlyGluPro...GlnAlaArgMetArgGluG 83
: : : : : : : : : : : : : :
461 GGGGCACACGACCCGCGCACACACGAGCGAGTTCAACACCGCCAGCGC... 506
83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99
|||: : : |||: : : |||: : : : :
507 .....CGGCTACCTGAACGCGCAGCTGCGCTACCAACACGTTGGG 545
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSe 112
: : : : : : : : : : : : : :
546 CACCGCGGCATCTACGAGCTGAAGCCACCCAGCTTCGTGCT..... 590
112 rThrAspProHisHisAspHis 121
: : : ||||| |||||
591 ....GAACAACGACACCATCGCCACCAT 614

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-35

seq_documentation_block:
; Sequence 35, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
```

US-08-470-566B-18

alignment\_scores:  
Quality: 87.00 Length: 126  
Ratio: 1.338 Gaps: 6  
Percent Similarity: 51.587 Percent Identity: 27.778

alignment\_block:

US-09-327-750D-31 x US-08-470-566B-18 ..

Align seg 1/1 to: US-08-470-566B-18 from: 1 to: 2004

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31

270 GAACGTGACATGGAGAGGTGATCTGTAGCCCAACGAGAACCTGAGCA 319

31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysV 48

320 ACAGCGTGGAGAGCCACTCGAGCACCACCACTGGAGCTACACCAACAC 365

48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64

366 .....CGAGGGCGCCAGCGTGGAGCGCGCATCGGTCCCAAGGGCATCAG 410

65 TyrArgTIPAsp..... 68

411 CTTCGGCGTGAGCTGAACCTACACGACGAGACCGTGGCCCGCAGGAGT 460

69 ....IleMetHisArgLeuGlyGluPro...GlnAlaArgMetArgGluG 83

461 GGGGCACACGACCGCGCACACAGCCAGTTCACACACCGCCAGCGC.... 506

83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99

507 .....CGGTACTCTGAACGCGCAACGTCGCTACACCAACGCTGGG 545

100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSe 112

546 CACCGCGCCACTACGACGTGAGCCACACCGCCACCGCTTCGTGCT..... 590

112 rThrAspProProHisHisAspHis 121

591 .....GAACAACGACACCATCGCCACCAT 614

seq\_name: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:US-08-469-334-18

seq\_documentation\_block:

Sequence 18, Application US/08469334

Patent No. 5990383

GENERAL INFORMATION:

APPLICANT: Warren, Gregory W

APPLICANT: Koziel, Michael G

APPLICANT: Mullins, Martha A

APPLICANT: Nye, Gordon J

APPLICANT: Carr, Brian

APPLICANT: Desai, Nalini M

APPLICANT: Kostichka, N. Kristy

APPLICANT: Duck, Nicholas B

APPLICANT: Estruch, Juan J

TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,334

FILING DATE: 06-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/463,483

FILING DATE:

APPLICATION NUMBER: US 08/314,594

FILING DATE: 09-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/037,057

FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Spruill, W. Murray

REGISTRATION NUMBER: 32,943

REFERENCE/DOCKET NUMBER: CGC 1695/CIP3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8615

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 2004 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1..2004

OTHER INFORMATION: /note= "Maize optimized DNA

OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"

US-08-469-334-1A

alignment\_scores:

Quality: 87.00 Length: 126

Ratio: 1.338 Gaps: 6

Percent Similarity: 51.587 Percent Identity: 27.778

alignment\_block:

US-09-327-750D-31 x US-08-469-334-18 ..

Align seg 1/1 to: US-08-469-334-18 from: 1 to: 1

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31

270 GAACGTGACATGGAGAGGTGATCTGTAGCCCAACGAGAACCTGAGCA 319

31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysV 48

320 ACAGCGTGGAGAGCCACTCGAGCACCACCACTGGAGCTACACCAACAC 365

48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64

366 .....CGAGGGCGCCAGCGTGGAGCGCGCATCGGTCCCAAGGGCATCAG 410

65 TyrArgTIPAsp..... 68

411 CTTCGGCGTGAGCTGAACCTACACGACGAGACCGTGGCCCGCAGGAGT 460

69 ....IleMetHisArgLeuGlyGluPro...GlnAlaArgMetArgGluG 83

461 GGGGCACACGACCGCGCACACAGCCAGTTCACACACCGCCAGCGC.... 506

83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99

507 .....CGGTACTCTGAACGCGCAACGTCGCTACACCAACGCTGGG 545

```

; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 08/463,483
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; FILING DATE: 09-SEP-1994
; APPLICATION NUMBER: US 08/314,594
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLv4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPLA(a) 80 kd protein from AB78"
;
US-08-471-046A-18

alignment_scores:
  Quality: 87.00 Length: 126
  Ratio: 1.338 Gaps: 6
Percent Similarity: 51.587 Percent Identity: 27.778

alignment_block:
US-09-327-750D-31 x US-08-471-046A-18 ..
Align seg 1/1 to: US-08-471-046A-18 from: 1 to: 2004
15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31
||||: :: |||:||||: :: |||:||||:
270 GAACGTGAGCATGGAGAAGGTGATCTGTGAGCCCAACGAGACCTGAGCA 319
31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysv 48
||||:||||:||||:||||: |||:||||: |||:||||:
320 ACAGCGTGGAGCGCACTCGACGACCACTGGAGCTACACCAACAC.... 365
48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64
||||:||||:||||: ||| ||| ||| |||
366 .....CGAGGCGCCAGCGTGGAGCGCGGATCGTCCCAAGGCATCAG 410
65 TyrArgTrpAsp.....
||| |||
411 CTTCGGCGTGCAGTGTAACCTACCAGCACGAGACCGTGGCCCGAGGAGT 460
69 .....IleMetHisArgLeuGlyGluPro....GlnAlaArgMetArgGluG 83
||||:||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
461 GGGCACACGACCGCGCAACACGACGAGTTCACACCGCCAGCGC.... 506
83 luAsnMetGluGluIleGlyGlnValArgGlnLeuMetGluLysLeu 99
||||: |||: |||: ||| ||| ||| ||| ||| ||| ||| ||| |||
507 .....CGGCTACCTGAAGCGCAACGCTCGCTACACCAACGTTGGG 545
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValse 112
```

```

546 CACCGCGGCATCTACGAGCTGAAGCCACCCAGCTTGCTGTCT..... 590
||||:||||:||||:
112 rthrAspProHisHisAspHis 121
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
591 ....GAACAACGACACCATGCGCCACCAT 614
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-566B-18

seq_documentation_block:
; Sequence 18, Application US/08470566B
; Patent No. 5872212
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5872212artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,566B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLv4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPLA(a) 80 kd protein from AB78"
```

546 CACCGGGCCATCTAGCAGCTGAAGCCACCCAGCTTGTGCT..... 590

112 rThrAspProHisHisAspHis 121

591 ....GAAACAAGACACCATCGCGCCACCAT 614

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-463-483A-18

seq\_documentation\_block:

; Sequence 18, Application US/08463483A

; Patent No. 5849870

; GENERAL INFORMATION:

; APPLICANT: Warren, Gregory W

; APPLICANT: Koziel, Michael G

; APPLICANT: Mullins, Martha A

; APPLICANT: Nye, Gordon J

; APPLICANT: Carr, Brian

; APPLICANT: Desai, Nalini M

; APPLICANT: Kostichka, N. Kristy

; APPLICANT: Duck, Nicholas B

; APPLICANT: Estruch, Juan J

; TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

; STATE: NY

; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30B

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/463.483A

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/314,594

; FILING DATE: 09-SEP-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/218,018

; FILING DATE: 23-MAR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/037,057

; FILING DATE: 25-MAR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Spruill, W. Murray

; REGISTRATION NUMBER: 32,943

; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-541-8615

; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2004 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1..2004

; OTHER INFORMATION: /note= "Maize optimized DNA

; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"

US-08-463-483A-18

alignment\_scores:

Quality: 87.00 Length: 126

Ratio: 1.338 Gaps: 6

Percent Similarity: 51.587 Percent Identity: 27.778

alignment\_block:

US-09-327-750D-31 x US-08-463-483A-18 ..

Align seg 1/1 to: US-08-463-483A-18 from: 1 to: 2004

15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31

||||: ||: ||||| ||: ||||| |||

270 GAACGTGAGCATGGAGAGGTGATCTGTAGCCCAACGAGAACCTGAGCA 319

||||: ||: ||||| ||: ||||| |||

31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysV 48

||||: ||: ||||| ||: ||||| |||

320 ACAGCGTGGAGGCCACTCGAGCACCACCTGGAGCTACACCAACAC... 365

||||: ||: ||||| ||: ||||| |||

48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64

||||: ||: ||||| ||: ||||| |||

366 .....CGAGGGCCAGCGTGGAGCGCGCATCGGTCCCAAGGCATCAG 410

||||: ||: ||||| ||: ||||| |||

65 TyrArgTrpAsp..... 68

||| ||: ||||| |||||

411 CTTCCGCGTGAGCGTGAACTACCCAGCAGCCGCGTGGCCAGGAGT 460

||||: ||: ||||| ||: ||||| |||

69 .....IleMetHisArgLeuGlyGluPro...GlnAlaArgMetArgGluG 83

||||: ||: ||||| ||: ||||| |||

461 GGGGCACACCGCCGCAACACCGCCAGTTCACACCGCCGCGC... 506

||||: ||: ||||| ||: ||||| |||

83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99

||||: ||: ||||| ||: ||||| |||

507 .....CGGTACCTGAACGCCAACGTCGCTACACCAACGCTGGG 545

||||: ||: ||||| ||: ||||| |||

100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValSe 112

||||: ||: ||||| ||: ||||| |||

546 CACCGCGCCATCTACGAGCTGAAGCCACCCAGCCAGCTTCGTGCT.... 590

||||: ||: ||||| ||: ||||| |||

112 rThrAspProHisHisAspHis 121

||||: ||: ||||| |||||

591 ....GAAACAAGACACCATCGCCACCAT 614

||||: ||: ||||| |||||

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-471-046A-18

seq\_documentation\_block:

; Sequence 18, Application US/08471046A

; Patent No. 5866326

; GENERAL INFORMATION:

; APPLICANT: Warren, Gregory W

; APPLICANT: Koziel, Michael G

; APPLICANT: Mullins, Martha A

; APPLICANT: Nye, Gordon J

; APPLICANT: Carr, Brian

; APPLICANT: Desai, Nalini M

; APPLICANT: Kostichka, N. Kristy

; APPLICANT: Duck, Nicholas B

; APPLICANT: Estruch, Juan J

; TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5866326artis Corporation

; STREET: 3054 Cornwallis Road

; CITY: Research Triangle Park

; STATE: NC

; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30B

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/471.046A



Quality: 87.00 Length: 126  
Ratio: 1.338 Gaps: 6  
Percent Similarity: 51.587 Percent Identity: 27.778

## alignment\_block:

US-09-327-750D-31 x US-08-471-033-18 ..  
Align seg 1/1 to: US-08-471-033-18 from: 1 to: 2004  
15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31  
||||: : : |||:||||| : : : : : |||  
270 GAACGTGAGCATGGAGAGGTGATCTTGAGCCCAACGAGAACTGAGCA 319  
: : : : : |||:||||| : : : : : |||  
31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysv 48  
: : : : : |||:||||| : : : : : |||  
320 ACACGTGGAGAGCCACTCGAGCACCACTGGAGCTACACCAACAC.... 365  
48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64  
|||||: : : |||:||||| : : : : : |||  
366 .....CGAGGGCCAGCGTGGAGCGCGGCATCGGTCCCAAGGCATCAG 410  
65 TyrArgTrpAsp..... 68  
||| : : : |||:||||| : : : : : |||  
411 CTTCGGCGTGAGCGTGAACCTACCAAGCAGCGTGGCCAGAGT 460  
69 ....IleMetHisArgLeuGlyGluPro...GlnAlaArgMetArgGluG 83  
: : : |||:||||| : : : ||| : : : |||  
461 GGGGCCACCAACCGCGGACACACCGAGTTCACACCGCCAGCGC.... 506  
83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99  
||||: : : |||:||||| : : : : : |||  
507 .....CGGCTACCTGAACGCGGCGGCATCGGTCCCAAGGCATG 545  
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValse 112  
: : : : : |||:||||| : : : : : |||  
546 CACCGCGCGCATCTAGACGTGAAGCCACCAACCGTGGCT..... 590  
112 rThrAspProHisHisAspHis 121  
: : : |||:||||| : : : : : |||  
591 .....GAACAACGACACCATCGCCACCAT 614

seq\_name: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:US-08-471-044-18

## seq\_documentation\_block:

Sequence 18, Application US/08471044  
Patent No. 5840868  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Malini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.308  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,044  
FILING DATE: 06-JUN-1995

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2004 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..2004  
OTHER INFORMATION: /note= "Maize optimized DNA  
OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"  
US-08-471-044-18

## alignment\_scores:

Quality: 87.00 Length: 126  
Ratio: 1.338 Gaps: 6  
Percent Similarity: 51.587 Percent Identity: 27.778

## alignment\_block:

US-09-327-750D-31 x US-08-471-044-18 ..  
Align seg 1/1 to: US-08-471-044-18 from: 1 to: 2004  
15 GluAsnValAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAl 31  
||||: : : |||:||||| : : : : : |||  
270 GAACGTGAGCATGGAGAGGTGATCTTGAGCCCAACGAGAACTGAGCA 319  
31 aAsnLysGlyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysv 48  
: : : : : |||:||||| : : : : : |||  
320 ACACGTGGAGAGCCACTCGAGCACCACTGGAGCTACACCAACAC.... 365  
48 alProArgGlyAsnArgArgPheArgValArgGlnProIleLeuGln 64  
|||||: : : |||:||||| : : : : : |||  
366 .....CGAGGGCCAGCGTGGAGCGCGGCATCGGTCCCAAGGCATCAG 410  
65 TyrArgTrpAsp..... 68  
||| : : : |||:||||| : : : : : |||  
411 CTTCGGCGTGAGCGTGAACCTACCAAGCAGCGTGGCCAGAGT 460  
69 ....IleMetHisArgLeuGlyGluPro...GlnAlaArgMetArgGluG 83  
: : : |||:||||| : : : ||| : : : |||  
461 GGGGCCACCAACCGCGGACACACCGAGTTCACACCGCCAGCGC.... 506  
83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99  
||||: : : |||:||||| : : : : : |||  
507 .....CGGCTACCTGAACGCGGCGGCATCGGTCCCAAGGCATG 545  
100 ArgGluLysGlnLeu.....SerHisSerLeuArgAlaValse 112  
: : : : : |||:||||| : : : : : |||

OM of: US-09-327-750D-31 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Mar 11, 2002 3:50 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=framet\_p2n\_model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09327750/runat\_11032002\_101154\_20340/app\_query.fasta\_1.1472  
-DB=Issued\_Patents\_NA -Qfmt=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09327750\_@CGN1\_1\_99 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-327-750D-31  
Query length: 128  
Database: Issued\_Patents\_NA:\*  
Database sequences: 351203  
Database length: 113238999  
Search time (sec): 146.090000

score_list:	Sequence	Strd Orig	zScore	EScore	Len	Documentation
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-044-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046B-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-334-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-300-529-18 +		87.00	158.56	0.3508	2004
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046B-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-334-35 +		87.00	155.90	0.4929	2576
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-044-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-044-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-566B-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-566B-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-334-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-469-334-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-300-529-17 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-300-529-26 +		87.00	155.58	0.5135	2655
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-033-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-044-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-483A-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-046A-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-566B-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-334-49 +		87.00	151.17	0.9042	4031
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-452-872-3 +		86.00	129.75	14.12	24979
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-452-872-3 +		86.00	129.75	14.12	24979
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-306-691B-23 +		79.00	139.85	3.86	2301
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-306-691B-23 +		79.00	139.85	3.86	2301
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US93-03985-3 +		79.00	139.85	3.86	2301
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US93-03985-3 +		79.00	139.85	3.86	2301
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US93-06251-78 +		79.00	139.85	3.86	2301
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US93-06251-78 +		79.00	139.85	3.86	2301

/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-513-974B-31 - 78.50 146.47 1.65  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-776-971-26 - 78.50 146.47 1.65  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-513-974B-322 - 78.50 144.55 2.11  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-776-971-103 - 78.50 144.55 2.11  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-998-416-141 - 78.00 150.69 0.9614  
seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-471-033-18

seq\_documentation\_block:  
; Sequence 18, Application US/08471033  
; Patent No. 5770696  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Gregory W  
; APPLICANT: Kozziel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,033  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pace, Gary M.  
; REGISTRATION NUMBER: P-40,403  
; REFERENCE/DOCKET NUMBER: C0C 1695/CIP3/DIV7 - SOLV3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8582  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2004 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..2004  
; OTHER INFORMATION: /note= "Maize optimized DNA  
; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"

US-08-471-033-18  
alignment\_scores:

```
|||||
212 ATGGAGTCCAAAGAGGAGCGGTTAAACAATCTCATCGTGGAATGT 261
17 lAsnGlnGluAsnAspGluLysGluGlnValAlaAsnLysG 34
|||||
262 CAACCAAGGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 311
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
312 GGGAGCCCTTGGCCCTTACCTTTGAATGTTAGTGAATCTGTGTGCTAGA 361
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
362 GGAACCGTAGGCGGTTCCGCGTTAGGCAGCCCATCTGCAGTATAGATG 411
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
|||||
412 GGACATAATGCATAGGCTTGGAGAGCCACAGGCAAGGATGAGAGGAGA 461
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
462 ATATGAAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGG 511
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
512 GAAACAGCTGAGTCATAGTTTGGCGGAGTCAGCACCTGATCCCCCTCA 561
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
562 CCATGACCATCAGATGAGTTTGCCTTATGCCCT 595
```

seq\_name: gb\_est1:BE266012

seq\_documentation\_block:

LOCUS BE266012 610 bp mRNA EST 13-JUL-2000  
DEFINITION 601194731F1 NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:3538735 5',  
mRNA sequence.

ACCESSION BE266012.1 GI:9139583

VERSION BE266012.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 610)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Plate: L1CM224 row: m column: 08

High quality sequence stop: 607.

FEATURES

source

1..610  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3538735"  
/clone\_lib="NIH\_MGC\_7"  
/tissue\_type="small cell carcinoma"  
/cell\_line="MGC3"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

ORIGIN

alignment\_scores:  
Quality: 663.00 Length: 129  
Ratio: 5.180 Gaps: 1  
Percent Similarity: 99.225 Percent Identity: 98.450  
alignment\_block:  
US-09-327-750D-31 x BE266012 ..  
Align seg 1/1 to: BE266012 from: 1 to: 610  
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17  
|||||  
172 ATGGAGTCCAAAGAGGAGCGGTTAAACAATCTCATCGTGGAATGT 221  
17 lAsnGlnGluAsnAspGluLysGluGlnValAlaAsnLysG 34  
|||||  
222 CAACCAAGGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 271  
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
|||||  
272 GGGAGCCCTTGGCCCTTACCTTTGAATGTTAGTGAATCTGTGTGCTAGA 321  
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67  
|||||  
322 GGAACCGTAGGCGGTTCCGCGTTAGGCAGCCCATCTGCAGTATAGATG 371  
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84  
|||||  
372 GGACATAATGCATAGGCTTGGAGAGCCACAGGCAAGGATGAGAGGAGA 421  
84 snMetGlu.ArgIleGlyGluGluValArgGlnLeuMetGluLysLeuAr 100  
|||||  
422 ATATGCAACAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGCTGAG 471  
100 gGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117  
|||||  
472 GGAACAGCAGTTGAGTCATAGTCTGCGGCGAGTCAGCACCTGATCCCCCTC 521  
117 ishHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||  
522 ACCATGACCATCAGATGAGTTTGGCCTTATGCCC 556

```

VERSION      AI291126.1  GI:39333900
KEYWORDS     EST.
SOURCE       Homo sapiens
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 669)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
             Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             DNA Library Arrayed by: Greg Lennon, Ph.D.
             Cloning Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 749 Std Error: 0.00
             Seq primer: -400P from Gibco
             High quality sequence stop: 466.
FEATURES     Location/Qualifiers
             1..669
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:1881915"
             /clone_lib="NCI-CGAP_Lu5"
             /tissue_type="carcinoid"
             /lab_host="DH10B"
             /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
             modified polylinker; 1st strand cDNA was prepared from
             neuroendocrine lung carcinoid, and was then primed with a
             Not I - oligo(dT) primer. Double-stranded cDNA was ligated
             into Eco RI adaptors (Pharmacia), digested with Not I and
             cloned into the Not I and Eco RI sites of the modified
             pT73 vector. Library is normalized. Library was
             constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT   186 a 132 c 185 g 164 t 2 others
ORIGIN
alignment_scores:
  Quality: 664.00      Length: 128
  Ratio: 5.270         Gaps: 0
  Percent Similarity: 98.438  Percent Identity: 98.438
alignment_block:
US-09-327-750D-31 x AI291126
Align seg 1/1 to: AI291126 from: 1 to: 669
1 MetGluSerLysGluGluArgAlaLeuAsnLeuIleValGluAsnVa 17
83 ATGGAGTCCAAGACGACGCGTTAAACATCTCATCTCGTGGAAATGT 132
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
133 CAACACGAAAAATGATGAAAAAGATGAAAAAGGCAAGTTGCTAATAAAG 182
34 lYcLupProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
183 GGGAGCCCTTGGCCCTTACCTTGAATGTTAGTGAATCTGTGTGCTAGTA 232
51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
233 GGAAACCGTAGCGGTTCGCGCTAGGCAGCCCATCTGTCAGTAGATG 282
67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGluA 84
283 GGACATAATGCTAGCTTGGAGAGCCACAGGCAAGGATGAGAGAGGAGA 332

```

```

84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100
|||||
333 ATATGGAAGAGATTTCGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 382
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHl 117
|||||
383 GAAACAGCTTGAGTCATAGTCTGCGGCGAGTCAGCACATGATCCCCCTCA 432
|||||
117 SHISAspHisAspGluPheCysLeuMetPro 128
|||||
433 CCATGACCATCAGATGAGTTTGCCTTATGCC 466
seq_name: gb_est2:BF237433

seq_documentation_block:
LOCUS      BF237433          711 bp      mRNA          EST          14-NOV-2000
DEFINITION 601842108F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4079856 5',
             mRNA sequence.
ACCESSION  BF237433
VERSION    BF237433.1  GI:11151351
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 711)
AUTHORS    NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LICM938 row: P column: 01
            High quality sequence stop: 708.
FEATURES   Location/Qualifiers
            1..711
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4079856"
            /clone_lib="NIH_MGC_46"
            /tissue_type="leiomyosarcoma cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Size-selected >500bp for average insert size
            1.8kb. Library constructed by Ling Hong in the laboratory
            of Gerald M. Rubin (University of California, Berkeley)
            using ZAP-cDNA synthesis kit (Stratagene) and Superscript
            II RT (Life Technologies). Note: this is a NIH_MGC
            Library."
BASE COUNT 184 a 152 c 218 g 157 t
ORIGIN
alignment_scores:
  Quality: 664.00      Length: 128
  Ratio: 5.228         Gaps: 0
  Percent Similarity: 99.219  Percent Identity: 98.438
alignment_block:
US-09-327-750D-31 x BF237433
Align seg 1/1 to: BF237433 from: 1 to: 711
1 MetGluSerLysGluGluArgAlaLeuAsnLeuIleValGluAsnVa 17

```

alignment\_scores:  
 Quality: 668.00 Length: 128  
 Ratio: 5.260 Gaps: 0  
 Percent Similarity: 99.219 Percent Identity: 99.219

alignment\_block:  
 US-09-327-750D-31 x BG705843 ..  
 Align seg 1/1 to: BG705843 from: 1 to: 813.

```

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
218 ATGGAGTCCAAAGAGAGAGAGCGGTAAACAAATCTCATCGTGGAAAATGT 267

17 lAsnGlnGluAsnAspGluLysGluGlnValAlaAlaAsnLysG 34
|||||
268 CAACCAAGGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 317

34 lyluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
318 GGGAGCCCTGGCCCTACCTTGAATGTTAGTGAATCTGTGCGCTAGA 367

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
368 GGAACCGTAGGCGGTTCGCGCTTAGGCAGCCCATCTGCAGTATAGATG 417

67 pAspIleMetHisArgLeuGluProGlnAlaArgMetArgGluGluA 84
|||||
418 GGACATAATGATAGCTTTGGAGAGCCACAGGCAAGGATGAGAGAGAGA 467

84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
468 ATATGGAAGAGTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 517

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
518 GAAAGCAGTTGAGTCATAGTCTGCGGCGAGTCAGCACTGATCCCTCA 567

117 sHisAspHisAspGluPheCysLeuMetPro 128
|||||
568 CCATGACCATCAGATGAGTTTGGCTTATGCCCC 601

```

seq\_name: gb\_estl:AI291270

seq\_documentation\_block:  
 LOCUS AI291270 511 bp mRNA EST 29-JAN-1999  
 DEFINITION qm16e06.x1 NCI\_CGAP\_Lu5 Homo sapiens cDNA clone IMAGE:1882018 3' similar to SW:HG74\_HUMAN Q00994 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74, mRNA sequence.

ACCESSION AI291270  
 VERSION AI291270.1 GI:3934044  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 511)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 Unpublished (1997)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)

Insert Length: 742 Std Error: 0.00  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 475.

#### FEATURES

Location/Qualifiers  
 1..511  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1882018"  
 /clone\_lib="NCI\_CGAP\_Lu5"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia) digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 152 a 103 c 155 g 100 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 666.00 Length: 128  
 Ratio: 5.286 Gaps: 0  
 Percent Similarity: 98.438 Percent Identity: 98.438

alignment\_block:

US-09-327-750D-31 x AI291270 ..

Align seg 1/1 to: AI291270 from: 1 to: 511

```

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
83 ATGGAGTCCAAAGAGAGAGCGGTAAACAAATCTCATCGTGGAAAATGT 132

17 lAsnGlnGluAsnAspGluLysGluGlnValAlaAlaAsnLysG 34
|||||
133 CAACCAAGGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 182

34 lyluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
183 GGGAGCCCTGGCCCTACCTTGAATGTTAGTGAATCTGTGCGCTAGA 232

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
233 GGAACCGTAGGCGGTTCGCGCTTAGGCAGCCCATCTGCAGTATAGATG 282

67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGluA 84
|||||
283 GGACATAATGATAGCTTTGGAGAGCCACAGGCAAGGATGAGAGAGAGA 332

84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
333 ATATGGAAGAGTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTCAGG 382

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
383 GAAAGCAGTGTAGTCATAGTCTGCGGCGAGTCAGCACTGATCCCTCA 432

117 sHisAspHisAspGluPheCysLeuMetPro 128
|||||
433 CCATGACCATCAGATGAGTTTGGCTTATGCCCC 466

```

seq\_name: gb\_estl:AI291126

seq\_documentation\_block:  
 LOCUS AI291126 669 bp mRNA EST 29-JAN-1999  
 DEFINITION qm15f02.x1 NCI\_CGAP\_Lu5 Homo sapiens cDNA clone IMAGE:1881915 3' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN); mRNA sequence.  
 ACCESSION AI291126

277 GAAAGCAGTTGAGTCATAGTCTGCGGCGAGTCAGCACTGATCCGCCCTCA 228

117 shIsAspHIsAspGluPheCysLeuMet 127

227 CCATGACCATCAGTGGTTCCTTATG 197

seq\_name: gb\_est2:BG707734

seq\_documentation\_block: 813 bp mRNA EST 07-MAY-2001  
LOCUS BG707734 60267125F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4794061 5',  
DEFINITION mRNA sequence.

ACCESSION BG707734.1 GI:13984377

VERSION BG707734

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 813)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cga@rs-re@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM10674 row: n column: 14

High quality sequence stop: 803.

Location/Qualifiers

1..813

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4794061"

/tissue\_type="hypothalmus"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified

pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag

); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',

size-selected for average insert size 2.3 kb and

normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NHGRI, National

Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 213 a 171 c 241 g 191 t 1 others

ORIGIN

alignment\_scores:

Quality: 672.00 Length: 128

Ratio: 5.250 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.219

alignment\_block:

US-09-327-750D-31 x BG707734 ..

Align seg 1/1 to: BG707734 from: 1 to: 817

1 MetGluSerLysGluAlaLeuAsnValGluAsnVa 17

|||||

221 ATGGAGTCCAAAGACGACGCGTAAACATCTCATCGTGGAAATGT 270

|||||

17 lasnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34

|||||

271 CAACAGGAAATGATGAAAGATGAAAGGAGCAGTCTGCTAATAAG 320

34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProAtg 50

|||||

321 GGGAGCCCTTGGCCCTTACCTTTGAATGTTAGTGAATCTGTGTGCTAGA 370

|||||

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67

|||||

371 GGAACCGTAGCGGTTCGCGTTAGGCAGCCCATCTGCAGTATAGATG 420

|||||

67 pspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84

|||||

421 GGACATAAATGCATAGCTTGGAGAGCCACAGGCAAGGATGAGAGAGAGA 470

|||||

84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100

|||||

471 ATATGGAAGAGATTGGGAGGAGGTGAGACGCTGATGGAAGCTGAGG 520

|||||

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117

|||||

521 GAAAGCAGCTTGCATAGTCTGCGGCGAGTCAGCACTGATCCGCCCTCA 570

|||||

117 shIsAspHIsAspGluPheCysLeuMetPro 128

|||||

571 CCATGACAAATCAGTGGTTCCTTATGCCC 604

|||||

seq\_name: gb\_est2:BG705843

seq\_documentation\_block: 813 bp mRNA EST 07-MAY-2001

LOCUS BG705843 602669329F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4792038 5',

DEFINITION mRNA sequence.

ACCESSION BG705843.1 GI:13980593

VERSION BG705843

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 813)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cga@rs-re@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM10669 row: j column: 07

High quality sequence stop: 772.

Location/Qualifiers

1..813

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4792038"

/tissue\_type="hypothalmus"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified

pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag

); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',

size-selected for average insert size 2.3 kb and

normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NHGRI, National

Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 207 a 168 c 243 g 194 t 1 others

ORIGIN

```

/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      181 a 159 c 226 g 160 t
ORIGIN

```

```

alignment_scores:
  Quality: 672.00      Length: 127
  Ratio: 5.291         Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-327-750D-31 x BG714974
..
Align seg 1/1 to: BG714974 from: 1 to: 726
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuLeuValGluAsnVa 17
|||||
219 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCGTGGAAATGT 268
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
269 CAACCAAGAAATCATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 318
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
319 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTCGCCCTAGA 368
51 GlyAsnArgArgPheArgValArgGlnProLeuLeuGlnTyrArgTr 67
|||||
369 GGAACCGTAGGGGTTCCCGTTAGGCAGCCCATCTCGCAGTATAGATG 418
67 pAspIleMethHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 84
|||||
419 GGACATAATGCATAGGCTTGAGAGCCACAGGCAAGGATGAGAGGAGA 468
84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100
|||||
469 ATATGGAAGAGGATGGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 518
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
519 GAAAGACAGTTGAGTCATAGTCTGCGGGCAGCTAGCACTGATCCCCCTCA 568
117 sHisAspHisAspGluPheCysLeuMet 127
|||||
569 CCATGACCATCAGCATGAGTTTGGCCTATG 599

```

seq\_name: gb\_est2:BF967675

```

seq_documentation_block:
LOCUS      BF967675      793 bp      mRNA      EST      23-JAN-2001
DEFINITION 602287383T1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4374549 3',
            mRNA sequence.
ACCESSION  BF967675
VERSION    BF967675.2 GI:12388141
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 793)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
```

```

TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    On Jan 16, 2001 this sequence version replaced gi:12334890.
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10038 row: f column: 22
High quality sequence start: 20
High quality sequence stop: 762.

```

FEATURES

```

1..793
Location/Qualifiers
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:4374549"
  /clone_lib="NIH_MGC_96"
  /tissue_type="hypothalamus"
  /lab_host="DH10B"
  /note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      185 a 235 c 170 g 203 t
ORIGIN

```

alignment\_scores:

```

  Quality: 672.00      Length: 127
  Ratio: 5.291         Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment\_block:

US-09-327-750D-31 x BF967675/rev

Align seg 1/1 to reverse of: BF967675 from: 1 to: 793

```

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuLeuValGluAsnVa 17
|||||
577 ATGGAGTCCAAAGAGGAGGCGTTAAACAATCTCATCGTGGAAATGT 528
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
527 CAACCAAGAAATCATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 478
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
477 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTCGCCCTAGA 428
51 GlyAsnArgArgPheArgValArgGlnProLeuLeuGlnTyrArgTr 67
|||||
427 GGAACCGTAGGGGTTCCCGTTAGGCAGCCCATCTCGCAGTATAGATG 378
67 pAspIleMethHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 84
|||||
377 GGACATAATGCATAGGCTTGAGAGCCACAGGCAAGGATGAGAGGAGA 328
84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100
|||||
327 ATATGGAAGAGGATGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 278
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||

```





```
VERSION BG707398.1 GI:13983707
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10678 row: d column: 04
High quality sequence stop: 814.

FEATURES
source
Location/Qualifiers
1..835
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4795347"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 220 a 176 c 246 g 193 t
ORIGIN

alignment_scores
Quality: 679.00 Length: 128
Ratio: 5.305 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block
US-09-327-750D-31 x BG707398 ...
Align seg 1/1 to: BG707398 from: 1 to: 835

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
248 ATGGAGTCCAAAGAGAGGAGGAGCGGTTAAACAATCTCATCGTGAAAAATGT 297

17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysC 34
|||||
298 CAACCAAGGAAATGATGAAAAGATGAAAGGAGCAAGTTGCTAATAAAG 347

34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
348 GGGAGCCCTGGCCCTACCTTGAATGTTAGTCAATCTGTCGCCTAGA 397

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
398 GGAACCGTAGGCGGTCCCGCTAGGCAGCCCATCTCGACGTATAGATG 447

67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluGlu 84
|||||
448 GGACATAATGATAGCTTGGAGAGGCCACAGGCAAGGATGAGAGAGAGA 497
```

```
84 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 100
|||||
498 ATATGCAAGAGATTGGGAGGAGTGCAGACAGCTGATGGAAGCTGAGG 547

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProHis 117
|||||
548 GAAACACAGTTGAGTCATAGTCTGCGGCGAGTCAGCAGTATCCCTCA 597

117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
598 CCATGACCATCAGCATGAGTTTTCCTTATGCC 631

seq_name: gb_est2:BF038787
seq_documentation_block:
LOCUS BF038787 1042 bp mRNA EST 20-OCT-2000
DEFINITION 601462119P1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3865348 5',
mRNA sequence.
ACCESSION BF038787
VERSION BF038787.1 GI:10745987
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1042)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DPV
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9608 row: f column: 05
High quality sequence stop: 732.

FEATURES
source
Location/Qualifiers
1..1042
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3865348"
/clone_lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 308 a 200 c 296 g 237 t 1 others
ORIGIN

alignment_scores
Quality: 679.00 Length: 128
Ratio: 5.305 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block
US-09-327-750D-31 x BF038787 ...
Align seg 1/1 to: BF038787 from: 1 to: 1042

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
85 ATGGAGTCCAAAGAGAGGAGGAGCGGTTAAACAATCTCATCGTGAAAAATGT 134

17 lAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysC 34
|||||
135 CAACCAAGGAAATGATGAAAAGATGAAAAGAGCAAGTTGCTAATAAAG 184
```

BASE COUNT 226 a 139 c 222 g 185 t  
 ORIGIN

alignment\_scores:  
 Quality: 679.00 Length: 128  
 Ratio: 5.305 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-31 x BE889313 ..

Align seg 1/1 to: BE889313 from: 1 to: 772

```

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
125 ATGGAGTCCAAAGAGAGACGCGTTAAACAATCTCATCTGGAATAATGT 174
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
175 CAACACGAGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTATAAAG 224
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
225 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATCTGTGTGCCTAGA 274
51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
275 GGAACCCGTAGCGGCTTCGCGTTAGGCAGCCCATCTGCAGTATAGATG 324
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 84
325 GGACATAATGCATAGCTTTGGAGACCCACAGGCAAGGATGAGAGGAGA 374
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
375 ATATGGAAGAGGATGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAG 424
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
425 GAAAGACAGTTCAGTCATAGTTTGGGGCAGTCAGCACTGTATCCCTCA 474
117 shiAspHisHisAspGluPheCysLeuMetPro 128
475 CCATGACCATCAGTATGAGTTTGGCTTATGCCCC 508

```

seq\_name: gb\_est2:BG709503

seq\_documentation\_block:  
 LOCUS BG709503 810 bp mRNA EST 07-MAY-2001  
 DEFINITION 602674781F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4797213 5',  
 mRNA sequence.

ACCESSION BG709503  
 VERSION BG709503.1 GI:13987904  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 810)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov  
 Plate: LLAM10683 row: a column: 22  
 High quality sequence stop: 770.

FEATURES  
 source

1. .810  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4797213"  
 /clone\_lib="NIH\_MGC\_96"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcagag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTNN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 206 a 168 c 247 g 187 t  
 ORIGIN

alignment\_scores:  
 Quality: 679.00 Length: 128  
 Ratio: 5.305 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-31 x BG709503 ..

Align seg 1/1 to: BG709503 from: 1 to: 810

```

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
218 ATGGAGTCCAAAGAGAGACGCGTTAAACAATCTCATCTGGAATAATGT 267
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
268 CAACACGAGAAATGATGAAAAAGATGAGAGGAGCAAGTTGCTATAAAG 317
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
318 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATCTGTGTGCCTAGA 367
51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
368 GGAACCCGTAGCGGCTTCGCGTTAGGCAGCCCATCTGCAGTATAGATG 417
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 84
418 GGACATAATGTCATAGCTTTGGAGACCCACAGGCAAGGATGAGAGGAGA 467
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
468 ATATGGAAGAGGATGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 517
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
518 GAAAGACAGTTCAGTCATAGTTTGGGGCAGTCAGCACTGTATCCCTCA 567
117 shiAspHisHisAspGluPheCysLeuMetPro 128
568 CCATGACCATCAGTATGAGTTTGGCTTATGCCCC 601

```

seq\_name: gb\_est2:BG707398

seq\_documentation\_block:

LOCUS BG707398 835 bp mRNA EST 07-MAY-2001  
 DEFINITION 602672810F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4795347 5',  
 mRNA sequence.  
 ACCESSION BG707398

241 GGGAGCCCTGGCCCTACCTTTGAATGTTAGTCAATACTGTGTGCTCCTAGA 290

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67

291 GGAACCGTAGCGGGTTCCCGGTAGGACGCCATCCTCGCATATAGATG 340

67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 84

341 GGACATAATGCATAGCTTGGAGAGCCACAGGCAAGGATGAGAGGAGA 390

84 snMetGluArgIleGlyGluGlnValArgGlnLeuMetGluLysLeuArg 100

391 ATATGGAAGAGGATGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGG 440

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117

441 GAAAGCAGTTGAGTCATAGTTTGGGGCAGTCAGACATGATCCCTCA 490

117 shIsAspHisAspGluPheCysLeuMetPro 128

491 CCATGACCATCAGCATGAGCTTTTGCCTTATGCC 524

seq\_name: gb\_est1.AL520931

seq\_documentation\_block:

LOCUS AL520931 717 bp mRNA EST 13-FEB-2001  
DEFINITION AL520931 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CS0DB002YA15 3  
prime. mRNA sequence.

ACCESSION AL520931

VERSION AL520931.1 GI:12784424

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 717)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..717

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CS0DB002YA15"

/clone\_lib="LTI\_NFL004\_NBC2"

/sex="male"

/tissue\_type="neuroblastoma cells"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed

by Life Technologies. Contact : Feng Liang Life

Technologies, a division of Invitrogen 9800 Medical Center

Drive Rockville, Maryland 20850, USA Fax : (1) 301 610

8371 Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 157 a 217 c 153 g 188 t 2 others

ORIGIN

alignment\_scores:

Quality: 679.00 Length: 128

Ratio: 5.305 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-31 x AL520931/rev ..

Align seg 1/1 to reverse of: AL520931 from: 1 to: 717

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17

530 ATGGAGTCCAAAGAGGACGAGCGTTAAACAATCTCATCGTGGAATGT 481

17 LAsnGlnGluAsnAspGluLysAspGluGlnValAlaAsnLysG 34

480 CAACACAGGAAATGATGAAAAAGATGAAAGGAGCAAGTTCCTAATAAG 431

34 LylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50

430 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTCAATACTGTGTGCTCCTAGA 381

51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67

380 GGAACCGTAGCGGGTTCCCGTTAGGACGCCATCCTCGCATATAGATG 331

67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 84

330 GGACATAATGCATAGCTTGGAGAGCCACAGGCAAGGATGAGAGGAGA 281

84 snMetGluArgIleGlyGluGlnValArgGlnLeuMetGluLysLeuArg 100

280 ATATGGAAGAGGATGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGG 231

101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117

230 GAAAGCAGTTGAGTCATAGCTTGGGGCAGTCAGACATGATCCCTCA 181

117 shIsAspHisAspGluPheCysLeuMetPro 128

180 CCATGACCATCAGCATGAGCTTTTGCCTTATGCC 147

seq\_name: gb\_est2:BE889313

seq\_documentation\_block:

LOCUS BE889313 772 bp mRNA EST 20-OCT-2000  
DEFINITION 601513201F1 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:3914401 5',  
mRNA sequence.

ACCESSION BE889313

VERSION BE889313.1 GI:10346503

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 772)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Arrayed by: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LHAM9736 row: b column: 02

High quality sequence stop: 763.

Location/Qualifiers

1..772

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3914401"

/clone\_lib="NIH\_MGC\_71"

/tissue\_type="leiomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

OM of: US-09-327-750D-31 to: EST:\* out\_format : pfs  
Date: Mar 11, 2002 2:16 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODE=framed\_p2n.model -DEV=xlh  
-Q/Cgn2\_1/USPTO\_spool/US09327750/runat\_11032002\_101153\_20308/app\_query.fasta\_1.1472  
-DB=EST -Qfmt=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09327750\_CGNI\_1.5654  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-327-750D-31  
Query length: 128  
Database: EST:\*  
Database sequences: 11351937  
Database length: 1077921985  
Search time (sec): 4085.940000

score_list:	Strd	Orig	ZScore	EScore	Len	Documentation
9b_est1:BE314909	+	679.00	1289.59	1.1e-62	594	1
9b_est1:AL520931	+	679.00	1287.82	1.4e-62	717	1
9b_est1:BE889313	+	679.00	1287.12	1.6e-62	772	1
9b_est1:BG709503	+	679.00	1286.67	1.7e-62	810	1
9b_est1:BG707398	+	679.00	1286.38	1.7e-62	835	1
9b_est1:BF038787	+	679.00	1284.29	2.3e-62	1042	1
9b_est1:BG285825	+	674.00	1275.65	6.9e-62	946	1
9b_est1:BG14974	+	672.00	1274.33	8.1e-62	726	1
9b_est1:BF97675	+	672.00	1273.96	9.0e-62	793	1
9b_est1:BG707734	+	672.00	1273.21	9.4e-62	817	1
9b_est1:BG705843	+	668.00	1265.62	2.5e-61	813	1
9b_est1:A1291270	+	666.00	1266.17	2.3e-61	511	1
9b_est1:A1291126	+	664.00	1259.81	5.2e-61	669	1
9b_est1:BF237433	+	664.00	1259.24	5.6e-61	711	1
9b_est1:BE266012	+	663.00	1258.77	6.0e-61	610	1
9b_est1:BE733217	+	661.00	1251.32	1.6e-60	897	1
9b_est1:BF967457	+	658.00	1247.04	2.7e-60	769	1
9b_est1:BG175745	+	658.00	1246.48	2.9e-60	816	1
9b_est1:BF685596	+	657.00	1244.55	3.7e-60	818	1
9b_est1:BG709294	+	646.00	1223.96	5.2e-59	782	1
9b_est1:AL520932	+	644.00	1219.53	9.2e-59	834	1
9b_est1:BG715353	+	638.00	1209.80	3.2e-58	694	1
9b_est1:BG3334975	+	636.00	1204.56	6.2e-58	807	1
9b_est1:BG655001	+	606.50	1151.87	5.4e-55	547	1
9b_est1:BG472167	+	606.50	1150.48	6.4e-55	634	1
9b_est1:BF971303	+	606.50	1150.13	6.7e-55	658	1
9b_est1:BG714823	+	606.50	1149.91	6.9e-55	674	1
9b_est1:AV702285	+	606.50	1149.68	7.1e-55	690	1
9b_est1:AV705808	+	606.50	1149.67	7.1e-55	691	1
9b_est1:AL526209	+	606.50	1149.62	7.2e-55	695	1
9b_est1:BG820179	+	606.50	1149.56	7.2e-55	699	1
9b_est1:BE526247	+	606.50	1148.49	8.3e-55	783	1
9b_est1:AV702643	+	606.50	1148.36	8.4e-55	794	1
9b_est1:AI929703	+	603.50	1144.05	1.5e-54	683	1
9b_est1:BG175659	+	599.50	1138.01	3.2e-54	576	1
9b_est1:AV732454	+	596.00	1135.73	4.3e-54	734	1
9b_est1:BG716327	+	596.00	1135.10	4.6e-54	386	1
9b_est1:BF16327	+	596.00	1126.82	1.3e-53	929	1
9b_est1:AI141583	+	592.00	1123.54	2.0e-53	585	1
9b_est1:BG709325	+	590.50	1117.40	4.5e-53	828	1

9b\_est1:AI826617 - 590.00 1119.74 3.3e-53 584 1 AI826617 wk35d03.x1 NCI\_CGAP  
9b\_est1:AW160832 + 584.50 1110.13 1.1e-52 531 1 AW160832 au76b12.y1 Schneide  
9b\_est1:BF346478 + 583.50 1106.22 1.9e-52 556 1 BF346478 602020342F1 NCI\_CGA  
9b\_est1:AV727777 + 580.50 1101.83 3.3e-52 569 1 AV727777 AV727777 HTC Homo s  
9b\_est1:AW953922 + 580.50 1101.23 3.6e-52 607 1 AW953922 EST365887 MAGE rese  
seq\_name: gb\_est1:BE314909  
seq\_documentation\_block:  
LOCUS BE314909 594 bp mRNA EST 26-OCT-2000  
DEFINITION 601140074F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3049290 5',  
mRNA sequence.  
ACCESSION BE314909  
VERSION BE314909.1 GI:9145006  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 594)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/BTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCM83 row: c column: 19  
High quality sequence stop: 594.  
FEATURES  
Location/Qualifiers  
source 1..594  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3049290"  
/clone\_lib="NIH\_MGC\_9"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: pOTB7; Site: 1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAGCAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 161 a 126 c 181 g 126 t  
ORIGIN  
alignment\_scores:  
Quality: 679.00 Length: 128  
Ratio: 5.305 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-327-750D-31 x BE314909 ..  
Align seg 1/1 to: BE314909 from: 1 to: 594  
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuLeuValGluAsnVa 17  
|||||  
141 ATGGAGTCCAAAGAGCAAGCGCGTAAACAATCTCATCTGGAATGCT 190  
17 IAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34  
|||||  
191 CAACGAGGAAATGATGAAAAAGATGAAAGGCAAGTTGCTATAAAG 240  
34 IyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
|||||



LOCUS DEFINITION	2901 bp	mrna	20-FEB-2001
ACCESSION	BC003254	clone MGC:5758, similar to dentatorubral pallidolusian atrophy,	
VERSION	BC003254.1	GI:13096906	
KEYWORDS	MGC.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
TITLE	1 (bases 1 to 2901)		
JOURNAL	Strausberg, R.		
	Direct Submission		
	Submitted (20-FEB-2001) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue procurement: Jeffrey Green M.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA sequencing by: Baylor College of Medicine Human Genome		
	Sequencing Center		
	Center code: BCM-HGSC		
	Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a>		
	Contact: vllallan@bcm.tmc.edu.		
	Villallon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,		
	A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,		
	Muzny, D.M., Gibbs, R.A.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>		
source	Series: IRAK Plate: 9 Row: 0 Column: 6.		
	Location/Qualifiers		
	1. .2901		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="MGC:5758 IMAGE:3500522"		
	/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating		
	ductal carcinoma. 5 month old virgin mouse."		
	/clone_lib="NCI_CGAP_Mam6"		
	/lab_host="DH10B"		
	/note="Vector: pCMV-SPORT6"		
	755. .2359		
CDS	/codon_start=1		
	/product="Similar to dentatorubral pallidolusian atrophy"		
	/protein_id="AAH03254.1"		
	/db_xref="GI:13096907"		
	/translation="MVPQGLPTLPCGLPKPIHVHTLHPRAKHEPTPSHQSLQSP		
	PQPLFSPPPAAPTTPPTATQIQEPAEYEPPEPPAPSPPPKPVVDVPS		
	HASQSAFNNKHLDRGFCNSARDLYFVPLEGSKLAKRADLVKVRREAQRAREKE		
	RRERERKERERERELERSVKLAQEGRAVPCPSLGPVPHRPFEPGSAVATVP		
	YLGPTPALRTLSVARHVMSPGNRHPPYVPLGAVDGLLGYNNPALYSSDPAARE		
	REREARDRDLRKLPGCEVPKPESEPLHGVPGCLDFPHRGGLALQGGPGLHPFP		
	PHPSLGLPERERLAAQAPALRDMYSIAERLAAERHAERAAALNDPLARLQMLNLT		
	PHHGHSHSHLHLQODAIHAASASVHPLIDPLASGSHLITRIPYPAGTLPNLPLPH		
	PLHENEVLRLHQLFAAYPRDLPASLSAPMSAAHQLQAMHAQSAELQRLALEQQQWLHAH		
	HPLHVSVPLPAQEDYYSHLKKESDKPL"		
BASE COUNT	623 a	958 c	791 g 529 t
ORIGIN			
alignment_scores:			
Quality:	384.50	Length:	119
Ratio:	3.697	Gaps:	3
Percent Similarity:	87.395	Percent Identity:	65.546
alignment_block:			
US-09-327-750D-32	x	BC003254	..

```
alignment_scores:
  Quality: 446.50      Length: 129
  Ratio: 3.917        Gaps: 3
  Percent Similarity: 88.372      Percent Identity: 68.217

alignment_block:
US-09-327-750D-32 x AF097437
Align seg 1/1 to: AF097437 from: 1 to: 2369

1 MetGluSerlysluylusArgAlaValAsnSerLeuSerMetGluAsnAl 17
1268 ATGGAGTCCAAAGAT...CAAGCGTGAAATAATCTCAACATGAGAGATGA 1314
17 aAsnGlnGluAsnGluGluLysGlu.....GlnValAlaAsnLysG 31
1315 CCATCAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1364
31 lyGluPro...LeuAlaLeuProLeuAspAlaGlyGluTyrCysValPro 46
1365 GCGAGCCAGCTGTGGCCCTGACCTCCAGGCTGGCAAAATCTGCGACCT 1414
47 ArgGlyAsnArgArgArgPheProValArgGlnProIleLeuGlnTyrAr 63
1415 AGAGGAGGTGCGAGCGGTTCGCGGTTCGCGAGCCATCGCTCACTATAG 1464
63 gTrpAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluG 80
1465 ATGGGACCTGATCCAGAGGTTCGGAGGCCAGAGGAGAGATGAGAGAGG 1514
80 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 96
1515 AGAACGTACAGAGGTTTGGGGGTGTGTGAGACAGCTCATGGAGAGCTG 1564
97 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 113
1565 AGGGAAGGAGCGTGTGAGCACACGCTTCGGCGGTTCGCGGTTCGCGCC 1614
113 oHisHisAspHisAspGluPheCysLeuMetPro 125
1615 TCATCATGACCACCATGATGAGTTTGCCTCATGCCCC 1651

seq_name: gb_ro:AF097438

seq_documentation_block:
LOCUS AF097438 835 bp mRNA ROD 11-APR-1999
DEFINITION Mus musculus brain expressed x-linked protein 1 (Bex1) mRNA,
complete cds.
ACCESSION AF097438
VERSION AF097438.1 GI:4580589
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 835)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070
REMARK Erratum: [[published erratum appears in Hum Mol Genet 1999
May;8(5):943]]
2 (bases 1 to 835)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
Location/Qualifiers
source
1..835
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
```

```
|||||
18762 CCATGACCATCAGATGAGTTTGCCTTATGCC 18795

seq_name: gb_ro:AF097439

seq_documentation_block:
LOCUS AF097439 785 bp mRNA ROD 13-APR-1999
DEFINITION Mus musculus brain expressed X-linked protein 2 (Bex2) mRNA,
complete cds.
ACCESSION AF097439
VERSION AF097439.1 GI:4580591
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 785)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070
REMARK Erratum:[published erratum appears in Hum Mol Genet 1999
May;8(5):943]]
REFERENCE 2 (bases 1 to 785)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
source
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="dbEST:W48832"
/chromosome="X"
/map="near Pip"
/dev_stage="embryo; 15.5 dpc"
1..785
/gene="Bex2"
139..528
/gene="Bex2"
/product="brain expressed X-linked protein 2"
/protein_id="AAD24430.1"
/db_xref="GI:4580592"
/translation="MESKVGQVKNLNMNDHQKEEKPKQDASKRDPVALPPEA
GDYVPRGRRRVRQPIVHYRWDLMRVGEPQGRMRRENVQREGDVRQLMEKLRE
QLSHSLRAVSTDPHPHHDHDFCLMP"
BASE COUNT 212 a 163 c 235 g 175 t
ORIGIN

alignment_scores:
Quality: 496.00 Length: 129
Ratio: 4.168 Gaps: 2
Percent Similarity: 92.248 Percent Identity: 72.093

alignment_block:
US-09-327-750D-32 x AF097439 ..
Align seg 1/1 to: AF097439 from: 1 to: 785
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
139 ATGGAGTCCAAAGTGGAACAAGCGGTGAAAAATCTCAACATGGAGAATGA 188
17 aAsnGlnGluAsnGluLysGlu.....GlnValAlaAsnLysG 31
189 CCATCAGGAAAGGAGGAAAGGAAAGCAAGCCACAGGATGTCGCAAAA 238
31 lyGluProLeu...AlaLeuProLeuAspAlaGlyGluTyrCysValPro 46
:::|||||: |||||||:|||||:|||||:|||||:|||||:|||||
```

```
239 GGGATCCGATTGTGGCCCTGCCTTTTCAAAGCTGGAGACTACTACGTGCCT 288
47 ArgGlyAsnArgArgPheProValArgGlnProileLeuGlnTyrAr 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
289 AGAGGAGGTGGCAGCGGTTCGGGGTTCGCAGCCCATCGTCACACTACAG 338
63 gTrpAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluG 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
339 ATGGGACCTTCATGCATAGGTGGGGAGCCCGCCAGGAGGATGAGAGAGG 388
80 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 96
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
389 AGAACGTACAGAGGTTTGGGGATGATGTGACAGCATCATGGAGAAGCTG 438
97 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
439 AGGGAAGGACAGCTGAGCCACAGCCCTGGCGGGTGTAGCACATGACCCGC 488
113 oHisHisAspHisHisAspGluPheCysLeuMetPro 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
489 TCATCATGACCAACCATGATGAGTTTGCCTTATGCC 525
seq_name: gb_ro:AF097437
seq_documentation_block:
LOCUS AF097437 2269 bp DNA ROD 05-AUG-1999
DEFINITION Mus musculus Bex1 protein (Bex1) gene, complete cds.
ACCESSION AF097437
VERSION AF097437.1 GI:5702155
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 2269)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070
PUBMED 10072429
REFERENCE 2 (bases 1 to 2269)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4030, Australia
FEATURES
source
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"
/map="near Pip"
/tissue_type="liver"
/join(409..509,866..942,1263..1873)
/gene="Bex1"
/product="Bex1 protein"
409..1873
/gene="Bex1"
/note="expressed in brain; X-linked"
1268..1654
/gene="Bex1"
/codon_start=1
/product="Bex1 protein"
/protein_id="AAD47168.1"
/db_xref="GI:5702156"
/translation="MESKDGQVKNLNMNDHQKEEKPKQDTRIRREPAVLTSEAG
KNCAPGRGRRRVRQPIAHYRWDLMRVGEPQGRMRRENVQREGDVRQLMEKLRE
QLSHSLRAVSTDPHPHHDHDFCLMP"
BASE COUNT 563 a 554 c 705 g 447 t
ORIGIN
```



/note="L1ME2 repeat: matches 4631. .6155 of consensus"  
 34528. .34786  
 /note="L1ME3A repeat: matches 4404. .4664 of consensus"  
 34787. .35085  
 /note="Alusg repeat: matches 1. .300 of consensus"  
 35086. .35412

alignment\_scores:  
 Quality: 602.50 Length: 128  
 Ratio: 4.898 Gaps: 1  
 Percent Similarity: 96.094 Percent Identity: 89.062

alignment\_block:  
 US-09-327-750D-32 x AL133348/rev ..

Align seg 1/1 to reverse of: AL133348 from: 1 to: 40584

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
 |||||  
 31132 ATGGAGTCCAAAGAGGAAACGAGCGTTAAACAATCTCATCTGGAATGT 31083

17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31  
 :|||  
 31082 CAACCGAGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAG 31033

31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47  
 |||||  
 31032 GGGAGCCCTGGCCCTACCTTTGAATTTAGTGAATACTGTGTGCCTAGA 30983

48 GlyAsnArgArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64  
 |||||  
 30982 GGAACCGTAGGCGGTTCCCGCTTAGCGAGCCATCTCGCAGTATAGATG 30933

64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81  
 |||||  
 30932 GGACATATCATAGGCTTGAGAGCCACAGGCAAGATGAGAGAGAGA 30883

81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97  
 |||||  
 30882 ATATGAAAGGATTTGGGAGGAGTGAGACAGCTGATGGAAGAGCTCAGG 30833

98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114  
 |||||  
 30832 GAAAGCAGTTGAGTCATAGTTTTCGGGCGAGTCAGCACTGATCCCTCA 30783

114 sHisAspHisHisAspGluPheCysLeuMetPro 125  
 |||||  
 30782 CCATGACCATCAGATGAGTTTTCGGCTTATGCCCC 30749

seq\_name: gb\_htg:HSU80B1

seq\_documentation\_block:  
 LOCUS HSU80B1 41029 bp DNA HTG 10-JUL-2001  
 DEFINITION Homo sapiens chromosome X clone LLOXNC01-80B1 map q22.1-22.3, \*\*\*  
 SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.

ACCESSION AL022169  
 VERSION AL022169.3 GI:13276704  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_CANCELLED.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 41029)

Bird.C.  
 Direct Submission  
 Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk

On Mar 12, 2001 this sequence version replaced gi:12750927.  
 ----- Genome Center

Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk  
 ----- Project Information  
 Center project name: cu80B1  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator ABI; 19% of reads  
 Chemistry: Dye-terminator Big Dye; 80% of reads  
 Consensus quality: 41009 bases at least Q40  
 Consensus quality: 41023 bases at least Q30  
 Consensus quality: 41027 bases at least Q20  
 Insert size: 41029; sum-of-contigs  
 Quality coverage: 51751; 0.6% error; agarose-fp  
 coverage: 7.67x in Q20 bases; sum-of-contigs Quality  
 coverage: 7.67x in Q20 bases; agarose-fp

\*\*\*\*\* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

FEATURES  
 Location/Qualifiers

1..41029  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /map="q22.1-22.3"  
 /clone="LLOXNC01-80B1"  
 /clone\_lib="LLOXNC01"  
 1..41029

misc\_feature  
 1..41029  
 /note="assembly\_fragment:00890"  
 BASE COUNT 13249 a 7796 c 7781 g 12203 t  
 ORIGIN

alignment\_scores:  
 Quality: 602.50 Length: 128  
 Ratio: 4.898 Gaps: 1  
 Percent Similarity: 96.094 Percent Identity: 89.062

alignment\_block:

US-09-327-750D-32 x HSU80B1 ..

Align seg 1/1 to: HSU80B1 from: 1 to: 41029

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
 |||||  
 18412 ATGGAGTCCAAAGAGGAAACGAGCGTTAAACAATCTCATCTGGAATGT 18461

17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31  
 :|||  
 18462 CAACCGAGAAATGATGAAAAAGATGAAAGGAGCAAGTTGCTAATAAG 18511

31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47  
 |||||  
 18512 GGGAGCCCTGGCCCTACCTTTGAATTTAGTGAATACTGTGTGCCTAGA 18561

48 GlyAsnArgArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64  
 |||||  
 18562 GGAACCGTAGGCGGTTCCCGCTTAGCGAGCCATCTCGCAGTATAGATG 18611

64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81  
 |||||  
 18612 GGACATATCATAGGCTTGAGAGCCACAGGCAAGATGAGAGAGAGA 18661

81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97  
 |||||  
 18662 ATATGAAAGGATTTGGGAGGAGTGAGACAGCTGATGGAAGAGCTCAGG 18711

98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114  
 |||||  
 18712 GAAAGCAGTTGAGTCATAGTTTTCGGGCGAGTCAGCACTGATCCCTCA 18761

114 sHisAspHisHisAspGluPheCysLeuMetPro 125

The true left end of clone LLOXNC01-105G4 is at 40485 in this sequence. The true right end of clone LLOXNC01-177E8 is at 100 in this sequence.

FEATURES	Location/Qualifiers	Source
	1..40584	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="X"	
	/map="q21.32-22.1"	
	/clone="RP1-79p11"	
	/clone_lib="RPC1-1"	
	19..292	
repeat_region	/note="AluJo repeat: matches 3..275 of consensus"	
repeat_region	470..1284	
repeat_region	/note="L1MB5 repeat: matches 5318..6157 of consensus"	
repeat_region	1295..1898	
repeat_region	/note="L1PA15 repeat: matches 5556..6157 of consensus"	
repeat_region	1899..2345	
repeat_region	/note="L1PA15 repeat: matches 4922..5370 of consensus"	
repeat_region	2323..2698	
repeat_region	/note="L1MB5 repeat: matches 4960..5343 of consensus"	
repeat_region	2927..3063	
repeat_region	/note="MER4A repeat: matches 164..305 of consensus"	
repeat_region	3655..9795	
repeat_region	/note="L1PA4 repeat: matches 2..6142 of consensus"	
repeat_region	9866..9982	
repeat_region	/note="FLAMC repeat: matches 6..124 of consensus"	
repeat_region	9989..10254	
repeat_region	/note="L1 repeat: matches 4027..4293 of consensus"	
repeat_region	10253..10852	
repeat_region	/note="L1M4 repeat: matches 4755..5333 of consensus"	
repeat_region	10853..11260	
repeat_region	/note="MLT1B repeat: matches 3..390 of consensus"	
repeat_region	11261..11474	
repeat_region	/note="L1M4 repeat: matches 4538..4755 of consensus"	
repeat_region	11480..11563	
repeat_region	/note="MSTB repeat: matches 1..80 of consensus"	
repeat_region	11564..12475	
repeat_region	/note="L1 repeat: matches 3039..3971 of consensus"	
misc_feature	14453..14861	GSS: Em:AQ729215"
repeat_region	15578..15854	/note="match: GSS: Em:AQ729215"
misc_feature	/note="AluJo repeat: matches 6..289 of consensus"	
repeat_region	complement(16007..16556)	
repeat_region	/note="match: GSS: Em:AQ772810"	
repeat_region	16197..16569	
repeat_region	/note="L1M4 repeat: matches 4267..4627 of consensus"	
repeat_region	16629..16714	
repeat_region	/note="L1M4C repeat: matches 1578..1666 of consensus"	
repeat_region	16902..17133	
repeat_region	/note="L1M4 repeat: matches 0..242 of consensus"	
repeat_region	17536..17728	
repeat_region	/note="MIR repeat: matches 1..191 of consensus"	
repeat_region	17746..18323	
repeat_region	/note="L2 repeat: matches 1599..2175 of consensus"	
repeat_region	18706..18755	
repeat_region	/note="L1MB3 repeat: matches 6034..6083 of consensus"	
repeat_region	18756..19236	
repeat_region	/note="MER6B repeat: matches 1..486 of consensus"	
repeat_region	19237..21108	
repeat_region	/note="MER66-internal repeat: matches 4808..6676 of consensus"	
repeat_region	21109..21163	
repeat_region	/note="MLT2B repeat: matches 391..445 of consensus"	
repeat_region	21164..21555	
repeat_region	/note="MER66B repeat: matches 3..391 of consensus"	
repeat_region	21557..21662	
repeat_region	/note="L1MB3 repeat: matches 6080..6185 of consensus"	
repeat_region	21665..21761	
repeat_region	/note="L1M repeat: matches 5299..5392 of consensus"	
repeat_region	21762..21929	
repeat_region	/note="L1MB8 repeat: matches 5902..6063 of consensus"	
repeat_region	21930..22212	
	/note="AluSg repeat: matches 3..284 of consensus"	
	22213..22321	
	/note="L1MB8 repeat: matches 6063..6173 of consensus"	
	22333..22740	
	/note="L1MD2 repeat: matches 5867..6288 of consensus"	
	22744..23094	
	/note="L1MD repeat: matches 2..342 of consensus"	
	23218..23524	
	/note="AluSg repeat: matches 1..305 of consensus"	
	23603..23947	
	/note="L1M4 repeat: matches 4916..5184 of consensus"	
	23948..24218	
	/note="AluJo repeat: matches 1..282 of consensus"	
	24219..24510	
	/note="L1M4 repeat: matches 4601..4916 of consensus"	
	24508..24621	
	/note="L1M4 repeat: matches 4374..4488 of consensus"	
	24650..25077	
	/note="MLT1C repeat: matches 1..466 of consensus"	
	25078..25914	
	/note="L1M4 repeat: matches 3497..4346 of consensus"	
	25920..26741	
	/note="L1M4 repeat: matches 2202..3065 of consensus"	
	26762..26906	
	/note="L1PA7 repeat: matches 6001..6145 of consensus"	
	26907..27135	
	/note="L1PA5 repeat: matches 5674..5902 of consensus"	
	27266..27509	
	/note="L1M4 repeat: matches 2164..2080 of consensus"	
	27557..27769	
	/note="L1MEC repeat: matches 253..460 of consensus"	
	28120..28167	
	/note="L2 repeat: matches 2703..2750 of consensus"	
	29043..29349	
	/note="MER58B repeat: matches 1..341 of consensus"	
	30389..30452	
	/note="16 copies 4 mer tata 89% conserved"	
	complement(30502)	
	/gene="dJ79p11.1"	
	complement(join(30502..31137,31514..31589,32016..>32134))	
	/note="match: ESTs: Em:AA317587 Em:AA358632 Em:AI214048	
	Em:AI929470 Em:AA319216 Em:AA317625 Em:D81392 Em:AA317418	
	Em:AA366528 Em:C15186 Em:AI291270 Em:AA160660 Em:AA612605	
	Em:AW160832 Em:AA384405 Em:AA160657 Em:AA081744	
	Em:AI251126 Em:C15591 Em:AA384268 Em:AI929703 Em:AI302026	
	Em:W19547 Em:AI141727 Em:AA946933 Em:AI141583 Em:H61106	
	Em:W60581 Em:AA641652 Em:AA706545 Em:AA464771 Em:AA862327	
	Em:W68380 Em:AA641652 Em:AA384269 Em:AA527696 Em:W48778	
	Em:AA633989 Em:AI659486 Em:AW167301 Em:AI929428	
	Em:AW379411 Em:C15948 Em:AI745669 Em:AW024587 Em:AW379207;	
	match: proteins: Tr:Q9Y516 Tr:Q9Y517"	
	/evidence=not_experimental	
	/product="dJ79p11.1 (novel protein similar to mouse Bex2 (brain-expressed X-linked protein 2))"	
	complement(30502..32134)	
	/gene="dJ79p11.1"	
	30507..30988	
	/note="match: STS: Em:G24641; match: STS: Em:T86927"	
	30508..30852	
	/note="match: STS: Em:R43117; match: STS: Em:G23964"	
	complement(30518..30523)	
	/gene="dJ79p11.1"	
	30541..30922	
	/note="match: STS: Em:H61107"	
	31543..32262	
	/note="CpG island"	
	/evidence=not_experimental	
	31647..31730	
	/note="42 copies 2 mer cc 64% conserved"	
	complement(32655..33316)	
	/note="match: GSS: Em:AQ489303"	
	33027..34514	
	/note="match: GSS: Em:AQ489303"	
	33027..34514	

/partial  
/note="Alu repeat: matches 298. .108 of consensus"  
10593. .10688  
/partial  
/note="Alu repeat: matches 96. .1 of consensus"  
10698. .10781  
/note="L1 element fragment"  
11112. .11403  
/note="Alu repeat: matches 1. .308 of consensus"  
11922. .12004  
/note="L1 element fragment"  
12062. .12308  
/note="L1 element fragment"  
12423. .12707  
/partial  
/note="Alu repeat: matches 302. .1 of consensus"  
12710. .12916  
/note="L1 element fragment"  
12979. .13081  
/note="L1 element fragment"  
13473. .13532  
/note="MLT2A1 element fragment"  
15871. .15928  
/note="L1 element fragment"  
18180. .18417  
/note="L1 element fragment"  
18778. .19056  
/note="Alu repeat: matches 1. .308 of consensus"  
22279. .22350  
/note="L1 element fragment"  
22449. .22995  
/note="L1 element fragment"  
23239. .23374  
/note="L1 element fragment"  
23371. .23606  
/note="MLT1B element fragment"  
23395. .23541  
/note="MLT1A element fragment"  
23705. .23778  
/note="MLT1A element fragment"  
23805. .23960  
/note="L1 element fragment"  
24011. .24541  
/note="L1 element fragment"  
24638. .24770  
/partial  
/note="Alu repeat: matches 150. .1 of consensus"  
24829. .30975  
/note="L1 element fragment"  
BASE COUNT 9462 a 6147 c 6129 g 9583 t  
ORIGIN

alignment\_scores:  
Quality: 602.50 Length: 128  
Ratio: 4.898 Gaps: 1  
Percent Similarity: 96.094 Percent Identity: 89.062

alignment\_block:

US-09-327-750D-32 x HSV870H8 ..

Align seg 1/1 to: HSV870H8 from: 1 to: 31321

1 MetGluSerLysGluLysArgAlaValAspSerLeuSerMetGluAsnAl 17  
3508 ATGAGTGCCAAAGAGGAGGAGCGCTTAAACAATCTCATCTGCGAAATGT 3557  
17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31  
3558 CAACGAGAAATGATGAAAGATGAAAGGAGCAAGTGTCTATTAAG 3607  
31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47  
|||||

3608 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCTAGA 3657  
48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64  
3658 GGAACCGTAGGCGGTCCCGTTAGGACGCCCATCTCGCAGTATAGTG 3707  
64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu 81  
3708 GGACATAATGCATAGGCTTGGAGAGCCACAGGCAAGCATGAGAGAGAGA 3757  
81 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 97  
3758 ATATGGAAGCATTTGGGAGAGGTGAGACAGCTGATGGAAAGCTGAGG 3807  
98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114  
3808 GAAAGCAGTTGAGTCATAGTCTGCGGGCAGTCAGCACTGATCCCCCTCA 3857  
114 sHisAspHisAspGluPheCysLeuMetPro 125  
3858 CCATGACCATCAGCATGAGTTTTCCTTATGCC 3891  
seq\_name: gb\_pr:AL133348  
seq\_documentation\_block:  
LOCUS AL133348 40584 bp DNA PRI 06-MAR-2000  
DEFINITION Human DNA sequence from clone RPL-79P11 on chromosome Xq21.32-22.1.  
Contains the gene for a novel protein similar to mouse Bex2  
(brain-expressed x-linked protein 2), ESTs, STSs, GSSs and a  
putative CpG island, complete sequence.  
ACCESSION AL133348  
VERSION AL133348.8 GI:7076386  
KEYWORDS HTG; Bex2; CpG island.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Wilson, S.  
Direct Submission  
Submitted (02-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Feb 24, 2000 this sequence version replaced gi:6997869.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em., EMBL; SW., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information  
on the WormPEP database can be found at  
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome X, constructed by the Sanger Centre Chromosome X Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/ChrX  
RPL-79P11 is from the library RPCI-1 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/  
VECTOR: PCYPAC2  
IMPORTANT: This sequence is not the entire insert of clone  
RPL-79P11 it may be shorter because we only sequence overlapping  
sections once, or longer because we arrange for a small overlap  
between neighbouring submissions.

```

64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81
|||||
301 GGACATAATGCATAGCTTGGAGAGCCACAGGCAAGGATGAGAGAGAGA 350

81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97
|||||
351 ATATGGAAGAGATTGGGAGAGAGGTGAGACAGCTGATGGAAGAGCTGAG 400

98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114
|||||
401 GAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCACTGATCCCCCTCA 450
451 CCATGACCATCAGCATGAGTTTGCCTTATGCCC 484

```

seq\_name: gb\_pat:AX078272

```

seq_documentation_block:
LOCUS AX078272 898 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 76 from Patent WO0107471.
ACCESSION AX078272
VERSION AX078272.1 GI:13157963
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Hillman,J.L., Lai,P., Tang,Y.T., Yue,H., Au-Young,J., Bandman,O.,
Azimzai,Y., Yang,J., Lu,D.A., Baughn,M.R., Patterson,C. and Shah,P.
TITLE Cell cycle and proliferation proteins
JOURNAL Patent: WO 0107471-A 76 01-FEB-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..898
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="incyte ID No: 5664154CB1"
BASE COUNT 250 a 186 c 251 g 211 t
ORIGIN

```

```

alignment_scores:
Quality: 602.50 Length: 128
Ratio: 4.898 Gaps: 1
Percent Similarity: 96.094 Percent Identity: 89.062

```

alignment\_block:

US-09-327-750D-32 x AX078272 ..

Align seg 1/1 to: AX078272 from: 1 to: 898

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
254 ATGAGTCCAAAGAGGACGAGGCTTAACAATCTCATCTGTGGAATGT 303

17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
|||||
304 CAACCAAGAAATGATGAAAGAGATGAAAGGAGCAAGTTGCTTAATAAAG 353

31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
|||||
354 GGAGGCCCTTGGCCCTACCTTGAATGTTAGTGAATCTGTGTCCTAGA 403

48 GlyAsnArgArgPhePheProValArgGlnProIleLeuGlnTyrArgTr 64
|||||
404 GGAACCGTAGGCGGTTCCTCCGTTAGGCAGCCCATCTCGCAGTATAGATG 453

64 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 81
|||||
454 GGACATAATGCATAGCTTGGAGAGCCACAGGCAAGGATGAGAGAGAGA 503

```

```

81 snMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeuArg 97
|||||
504 ATATGGAAGAGATTGGGAGAGAGGTGAGACAGCTGATGGAAGAGCTGAG 553

98 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 114
|||||
554 GAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCACTGATCCCCCTCA 603

114 sHisAspHisHisAspGluPheCysLeuMetPro 125
|||||
604 CCATGACCATCAGCATGAGTTTGCCTTATGCCC 637

```

seq\_name: gb\_pr:HSV870H8

```

seq_documentation_block:
LOCUS HSV870H8 31321 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from cosmid V870H8, between markers DXS366 and
DXS87 on chromosome X contains ESTs.
ACCESSION Z70233
VERSION Z70233.1 GI:12335542
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Whiteley,M.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-1995) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone V870H8 is at 1 in this sequence. The
true left end of clone V693A8 is at 3121.
V870H8 is from the human chromosome X-specific cosmid library.
FEATURES
source
1..31321
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="GHC-870H8"
/clone_lib="SCCV"
475..701
/note="L1 element fragment"
802..988
/note="L1 element fragment"
2533..4138
/note="L1 element fragment"
4189..4246
/note="match: multiple ESTs"
7496..7874
/note="29 copies of 2 mer 91 & conserved"
8022..8108
/note="L1 element fragment"
8302..8493
/note="L1 element fragment"
8585..8671
/note="L1 element fragment"
8769..9010
/note="L1 element fragment"
9259..9384
/note="L1 element fragment"
9624..9731
/note="MLTIC element fragment"
9781..9948
/note="MLTIC element fragment"
9890..9948
/note="MLTIC element fragment"
10418..10592
/note="MLTIB element fragment"

```

```

seq_name: gb_pr:AF237783
seq_documentation_block:
LOCUS AF237783 791 bp mRNA PRI 02-SEP-2000
DEFINITION Homo sapiens brain-expressed protein BEX1 (BEX1) mRNA, complete
cds.
ACCESSION AF237783
VERSION AF237783.1 GI:9963898
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 791)
AUTHORS Yang, Q.S., Ying, K., Xie, Y., and Mao, Y.M.
TITLE A Novel Human X-linked Brain Expressed Protein
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 791)
AUTHORS Mao, Y.M., Xie, Y., Yang, Q.S., Wu, H., Lin, S. and Ying, K.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2000) Genetic Research Institute, Fudan
University, 220 Handan Rd., Shanghai 200433, P.R.China
FEATURES
source
1..791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="between DXS990 and DX1059 by RH (Stanford G3)"
1..791
/gene="BEX1"
172..549
/gene="BEX1"
/codon_start=1
/product="brain-expressed protein BEX1"
/protein_id="AAG09752.1"
/db_xref="GI:9963898"
/translation="MESKEKLAVNSLSMENANQENEEKEQVANKGEPLALPLDAGEYC
VPRGNRRFRVROPILQYRWDMHRLGEPQARMREENMERIGEVRQLMEKLRKQLS
HSLRAVSTDPHHDHDEFCLMP"
BASE COUNT 214 a 172 c 218 g 187 t
ORIGIN

alignment_scores:
Quality: 639.00 Length: 125
Ratio: 5.238 Gaps: 0
Percent Similarity: 97.600 Percent Identity: 96.800

alignment_block:
US-09-327-750D-32 x AF237783 ..
Align seg 1/1 to: AF237783 from: 1 to: 791

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
172 ATGAGTCCAAAGAGAACTAGCAGTAACAGCTCAGCATGCGAAATGC 221

17 aAsnGlnGluAsnGlnGluLysGluGlnValAlaAsnLysGlyClyProL 34
|||||
222 CAACCAAGAAATTAAGAAAGGAGCAAGTGTCTAATAAAGGGGAGCCCT 271

34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
272 TGGCCCTCCCTTGGATGCTGGTGAATCTGTGTGCTAGAGGAATCGT 321

51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
|||||
322 AGGCGGTTCGCGTGTAGGCACCCCATCTGCAGTATAGATGGGATATGAT 371

67 tHisArgLeuGlyClyProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
372 GCATAGGCTTGGAGAACACAGGCAAGGATGAGAGAGAAATATGAAA 421

```

---

```

84 rgileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
422 GGATGGGGAGGGGTGAGACAGCTGATGGAACCTGAGGGAACAG 471

101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117
|||||
472 TTGAGTCATAGTCTGCGGGCAGTCAGCACTGACCCCTCACCATGACCA 521

117 SHISAspGluPheCysLeuMetPro 125
|||||
522 TCATGATGAGTTTGGCTTATGCC 546

seq_name: gb_pr:AF251053
seq_documentation_block:
LOCUS AF251053 744 bp mRNA PRI 15-APR-2001
DEFINITION Homo sapiens X-linked protein mRNA, complete cds.
ACCESSION AF251053
VERSION AF251053.1 GI:13625167
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 744)
AUTHORS Mao, Y., Xie, Y., Zhou, Z., Zhao, W., Wang, W., Huang, Y.,
Wang, S., Tang, R., Chen, X. and Wu, C.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2000) Institute of Genetics, School of Life
Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R.
China
FEATURES
Location/Qualifiers
1..744
/organism="Homo sapiens"
/db_xref="taxon:9606"
101..487
/codon_start=1
/product="X-linked protein"
/protein_id="AAK34943.1"
/db_xref="GI:13625168"
/translation="MESKEERALNNLIVENQENDEKEQVANKGEPLALPLNVS
EYCVPRGNRRFRVROPILQYRWDMHRLGEPQARMREENMERIGEVRQLMEKLRK
QLSHSLRAVSTDPHHDHDEFCLMP"
BASE COUNT 222 a 134 c 208 g 180 t
ORIGIN

alignment_scores:
Quality: 602.50 Length: 128
Ratio: 4.898 Gaps: 1
Percent Similarity: 96.094 Percent Identity: 89.062

alignment_block:
US-09-327-750D-32 x AF251053 ..
Align seg 1/1 to: AF251053 from: 1 to: 744

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
101 ATGGAGTCCAAAGAGAAAGAGCGCTTAACAATCATCGTGGAAATGT 150

17 aAsnGlnGluAsn.....GluGluLysGluGlnValAlaAsnLysG 31
|||||
151 CAACCAAGAAATGATGAAAAAGATGAAAGGAGCAAGTGTCTAATAAAG 200

31 lyGluProLeuAlaLeuProLeuAspAlaGlyGluTyrCysValProArg 47
|||||
201 GGGAGCCCTTGGCCCTACCTTGAATGTTAGTGAATACTGTGTGCTAGA 250

48 GlyAsnArgArgPheProValArgGlnProIleLeuGlnTyrArgTr 64
|||||
251 GGAACCGTAGGCGGTTCCCGCTTAGCGCCATCTCGCAGTATAGATG 300

```

RP1-198P4 It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.  
The true left end of clone LLOXNC01-221F2 is at 35615 in this sequence. The true right end of clone LLOXNC01-73E8 is at 100 in this sequence.

## FEATURES

## source

## Location/Qualifiers

1..35714  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="q22"  
/clone="RP1-198P4"  
/clone.lib="RPCI-I"

repeat\_region  
1..258  
/note="LIP3 repeat: matches 5666. 5925 of consensus"  
repeat\_region  
256..537  
/note="LIP3 repeat: matches 5391. 5672 of consensus"  
repeat\_region  
1491..1749  
/note="AluSc repeat: matches 39. 297 of consensus"  
repeat\_region  
2182..3810  
/note="LIP3 repeat: matches 4533. 6150 of consensus"  
repeat\_region  
5169..5687  
/note="LIP3 repeat: matches 4798. 6308 of consensus"  
repeat\_region  
6976..7665  
/note="LIP3 repeat: matches 4113. 4798 of consensus"  
repeat\_region  
8039..9213  
/note="LIP3 repeat: matches 2945. 4113 of consensus"  
repeat\_region  
9523..10437  
/note="LIP3 repeat: matches 2016. 2945 of consensus"  
repeat\_region  
10736..11494  
/note="LIP3 repeat: matches 1262. 2016 of consensus"  
repeat\_region  
11807..12251  
/note="LIP3 repeat: matches 310. 755 of consensus"  
repeat\_region  
12259..12331  
/note="HERVFBH21 repeat: matches 5. 77 of consensus"  
repeat\_region  
12332..12796  
/note="LIP3 repeat: matches 1. 490 of consensus"  
repeat\_region  
16162..16371  
/note="LIP3 repeat: matches 1. 322 of consensus"  
repeat\_region  
17121..17424  
/note="LIP3 repeat: matches 1. 322 of consensus"  
repeat\_region  
18259..18290  
/note="LIP3 repeat: matches 1. 322 of consensus"  
repeat\_region  
18505..18672  
/note="LIP3 repeat: matches 1. 181 of consensus"  
repeat\_region  
19478..19814  
/note="LIP3 repeat: matches 1. 345 of consensus"  
repeat\_region  
21061..21101  
/note="LIP3 repeat: matches 1. 345 of consensus"  
repeat\_region  
21247..21781  
/note="LIP3 repeat: matches 42. 85 of consensus"  
misc\_feature  
/note="LIP3 repeat: matches 44. 568 of consensus"  
24465..24492  
/note="Single clone region"  
Single clone region  
24531..24562  
/note="Weak data"  
Weak data  
24563  
/note="Single clone region"  
Single clone region  
24711..24865  
/note="region covered by subclones from a PAC DNA PCR product only at 12x coverage."  
region covered by subclones from a PAC DNA PCR product only at 12x coverage.  
24758..25460  
/note="CpG island"  
evidence=not\_experimental  
24875..24940  
/note="33 copies 2 mer cc 66% conserved"  
26795..26926  
/note="66 copies 2 mer ct 63% conserved"

repeat\_region  
27088..27133  
/note="23 copies 2 mer tg 100% conserved"  
repeat\_region  
27461..27777  
/note="LIMB8 repeat: matches 5851. 6171 of consensus"  
repeat\_region  
27802..28232  
/note="LIMC/D repeat: matches 5192. 5403 of consensus"  
repeat\_region  
28536..28930  
/note="LIMC/D repeat: matches 4802. 5192 of consensus"  
repeat\_region  
29240..29341  
/note="TIGGER1 repeat: matches 484. 587 of consensus"  
repeat\_region  
29603..30740  
/note="TIGGER1 repeat: matches 587. 1757 of consensus"  
repeat\_region  
31027..31365  
/note="TIGGER1 repeat: matches 1757. 2105 of consensus"  
repeat\_region  
31664..31691  
/note="TIGGER1 repeat: matches 2105. 2132 of consensus"  
repeat\_region  
31809..32015  
/note="TIGGER1 repeat: matches 2132. 2246 of consensus"  
repeat\_region  
32323..32389  
/note="TIGGER1 repeat: matches 2246. 2313 of consensus"  
repeat\_region  
34353..34636  
/note="MLT1-INTERNAL repeat: matches 607. 889 of consensus"  
repeat\_region  
35149..35376  
/note="MLT1-INTERNAL repeat: matches 1001. 1233 of consensus"  
repeat\_region  
35523..35670  
/note="MLT1-INTERNAL repeat: matches 8. 159 of consensus"  
BASE COUNT 9708 a 7094 c 6805 g 12107 t  
ORIGIN  
  
alignment\_scores:  
Quality: 653.00 Length: 125  
Ratio: 5.266 Gaps: 0  
Percent Similarity: 99.200 Percent Identity: 98.400  
  
alignment\_block:  
US-09-327-750D-32 x HS198P4/rev ..  
  
Align seg 1/1 to reverse of: HS198P4 from: 1 to: 35714  
  
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
|||||  
24358 ATGGAGTCCAAAGAGAAACGACGAGTAAACAGTCTCAGCATGGAATGC 24309  
17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGlnProL 34  
|||||  
24308 CAACCAAGAAATGAAGAAAGGAGCAAGTTGCTAATAAGGGAGCCCT 24259  
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50  
|||||  
24258 TGGCCCTCCCTTTGGATGCTGTGTAATCTGTGTGCTAGAGGAATCGT 24209  
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67  
|||||  
24208 AGCGGTTCCGGCTTAGGCGAGCCCTCCTGCAGTATAGATGGATATGAT 24159  
67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluAsnMetGluA 84  
|||||  
24158 GCATAGGCTTGGAGAACACACAGCAGGATGAGAGAGAGATATCGAAA 24109  
84 rGleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100  
|||||  
24108 GGATTGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGGGAAGACAG 24059  
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117  
|||||  
24058 TTGAGTCATAGTCTGCGGCGAGTCAGCAGTACGCCCTCACCCTACCA 24009  
117 sHisAspGluPheCysLeuMetPro 125  
|||||  
24008 TCATGATGAGTTTGGCTTATGCCC 23984



332 GGAATATACAGGGGGGGGAGTAGGACGACTTGTCCCTAATTTTCGATG 381  
 67 pAspIleMetHisArg...LeuGlyGluProGlnAlaArgMetArgGluG 83  
 382 GGCCATACCTTAATAGGCATATTGAGCACAATGAGCGAGA.....G 422  
 83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99  
 423 ATGATGTAGAAAGGTTGTAGGGCAGATGGAATCAAGAGAAAGACT 472  
 100 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116  
 473 AGGGAACAGCAGATGAGGCACATATATCGCTTCCAAACTCCTGAACCT.. 520  
 116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128  
 521 .....GACAACCATATTGACTTTTGCCTCATACCT 550

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH13750

seq\_documentation\_block:

ID: AAH13750 standard; cDNA; 1229 BP.

AC AAH13750;

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:10656.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126..

XX 29-JUL-1999; 93JP-0248036.

XX 27-AUG-1999; 95JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs.

XX Claim 8: SEQ ID 10656; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 1229 BP; 351 A; 243 C; 296 G; 339 T; 0 other;

alignment\_scores:  
 Quality: 187.00 Length: 129  
 Ratio: 2.226 Gaps: 4  
 Percent Similarity: 65.116 Percent Identity: 37.984

alignment\_block:

US-09-327-750D-31 x AAH13750 ..

Align seg 1/1 to: AAH13750 from: 1 to: 1229

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17  
 191 ATGGAGTCCAAAGAGAACTAGCGCAACAATCTCAACGGGGAATGC 240  
 17 lAsnGlnGluAsnAspGluLysGluGlnValAlaAlaAsnLysG 34  
 241 CCAACAAGAAAC.....GAAGGAGGGGAGCAGCCGCCACCCAGA 281  
 34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
 282 ATCAACAAGAATCCCGCCATTTCGGAGGGGTGAAGCCCAAGACCTGGA 331  
 51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67  
 332 GGAATATACAGCGGGGGGAGTTAGCGACTTGTCCCTAATTTTCGATG 381  
 67 pAspIleMetHisArg...LeuGlyGluProGlnAlaArgMetArgGluG 83  
 382 GGCCATACCTTAATAGGCATATTGAGCACAATGAGCGAGA.....G 422  
 83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99  
 423 ATGATGTAGAAAGGTTGTAGGGCAGATGGAATCAAGAGAAAGACT 472  
 100 ArgGluLysGlnLeuSerHisSerHisLeuArgAlaValSerThrAspProPr 116  
 473 AGGGAACAGCAGATGAGGCACATATATCGCTTCCAAACTCCTGAACCT.. 520  
 116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128  
 521 .....GACAACCATATTGACTTTTGCCTCATACCT 550

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH85548

seq\_documentation\_block:

ID: AAC85548 standard; cDNA; 1364 BP.

XX AAC85548;

XX 04-JUN-2001 (first entry)

DE cDNA encoding CD1FF-4, Incyte ID No. 1990956CB1.

XX Human; cell differentiation; CD1FF; agonist; antagonist; epilepsy;  
 KW cell proliferation; Alzheimer's disease; schizophrenic disorder;  
 KW arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 250...612



```
441 GGCCATACCTAATAGGCATATTGAGCACAATGAGCGAGA.....G 481
83 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeu 99
482 ATGATGTAGAAAGGTTTGTAGGGCAGATGATGGAATCAAGAGAACT 531
100 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
532 AGGGAACAGCAGATGAGGCACATATATCGCTTCCAAACTCTGAACCT.. 579
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
580 .....GACACCAATTATGACTTTTGGCTCATACCT 609
```

```
FT /*tag= a
FT /product= "CDIFF-4"
XX WO200119860-A2.
XX 22-MAR-2001.
XX PF 14-SEP-2000; 2000WO-US25435.
XX PR 15-SEP-1999; 99US-0154140.
XX PR 06-DEC-1999; 99US-0169155.
XX PA (INCY-) INCYTE GENOMICS INC.
XX Tang YT, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;
XX Baughn MR, Lu DAM, Bandman O, Shih LL, Patterson C;
XX WPI: 2001-211447/21.
XX P-PSDB; AAB47126.
XX Isolated polypeptides and polynucleotides involved in cell
PT differentiation are used for treatment, prevention and diagnosis of
PT cell proliferative, developmental and neurological disorders e.g.
PT cancer and Alzheimer's disease -
XX Claim 5; Page 121; 137pp: English.
XX The sequences given in AAC85545-72 encode human polypeptides involved
CC in cell differentiation (CDIFF). CDIFF polypeptides and agonists of
CC these are used to treat a disease or condition associated with
CC decreased expression of functional CDIFF. An antagonist of CDIFF is
CC used to treat a disease or condition associated with over expression
CC of functional CDIFF. CDIFF polypeptides may be used for the treatment,
CC prevention and diagnosis of cell proliferative, developmental and
CC neurological disorders, such as Alzheimer's disease, schizophrenic
CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus
CC and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus
CC REX-3. This sequence maps to chromosome 1 within the interval from
CC 152.2 to 157.4 centiMorgans, to chromosome 3 within the interval from
CC 157.4 to 158.0 centiMorgans, and to the X chromosome within the interval
XX from 104.9 to 150.3 centiMorgans.
SQ Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;
```

```
alignment_scores:
Quality: 187.00 Length: 129
Ratio: 2.226 Gaps: 4
Percent Similarity: 65.116 Percent Identity: 37.984
alignment_block:
US-09-327-750D-31 x AAC85548
Align seg 1/1 to: AAC85548 from: 1 to: 1364
1 MetGluSerLysGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
250 ATGGAGTCCAAAGAGAACTAGCGGCAACAATCTCAACGGGGAATGC 299
17 lAsnGlnGluAsnAspGluLysGluGlnValAlaAsnLysG 34
300 CCAACAAGAAAC.....GAAGAGGGGAGCAGGCCCCACGCAGA 340
34 lYGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
341 ATGAAGAGAAATCCCGCATTTGGGGGGGTGAAGCCAGACCTTGA 390
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
391 GGAATATACGGCGGGGAGTTAGCGACTTGTCCTAATTTTCGATG 440
67 pAspIleMetHisArg...LeuGlyGluProGlnAlaArgMetArgGluG 83
```

**THIS PAGE BLANK (USPTO)**

OM of: US-09-327-750D-31 to: GenEmbl: \* out\_format : pfs  
Date: Mar 11, 2002 3:34 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame2.p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101154\_20319/app\_query.fasta\_1.1472  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOFCU=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MAPRIX=blsom62 -TRANS=human40.cdi  
-LIST=45 -DLOCALIGN=200 -THR SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09327750\_@CGN1\_1.8673  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-31  
Query length: 128  
Database: GenEmbl: \*  
Database sequences: 1472140  
Database length: 341344837  
Search time (sec): 4557.230000

score\_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
gb_pr:AF251053	+	679.00	1230.71	2.8e-60	744	AF251053 Homo sapiens X-linked
gb_pat:AX078272	+	679.00	1229.04	3.5e-60	898	AX078272 Sequence 76 from Patent
gb_pr:HSV870H8	-	679.00	1197.46	2.0e-58	31321	Z702333 Human DNA sequence fr
gb_pr:AL1133348	-	679.00	1195.16	2.7e-58	40584	AL1133348 Human DNA sequence fr
gb_hcg:HSU80B1	+	679.00	1195.02	2.7e-58	41029	AL022169 Homo sapiens chromos
gb_pr:AF183416	+	606.50	1099.72	5.6e-53	642	AF183416 Homo sapiens ovarian
gb_pr:AF220189	+	606.50	1097.45	7.5e-53	828	AF220189 Homo sapiens uncharac
gb_pr:HS198P4	-	606.50	1063.99	5.5e-51	35714	AL008708 Human DNA sequence f
gb_pr:AF237783	+	592.50	1072.31	1.9e-51	791	AF237783 Homo sapiens brain-exp
gb_to:AF097439	+	502.50	908.13	2.6e-42	785	AF097439 Mus musculus brain exp
gb_to:AF097437	+	452.00	806.54	1.2e-36	2269	AF097437 Mus musculus Bexl pro
gb_to:AF097438	+	450.00	811.78	6.1e-37	835	AF097438 Mus musculus brain exp
gb_to:AF051347	+	450.00	811.74	6.1e-37	838	AF051347 Mus musculus REX-3 mR
gb_to:BC003254	+	388.00	687.56	5.1e-30	2901	BC003254 Mus musculus, Similar
gb_scs:G24641	-	339.00	613.70	6.6e-26	504	G24641 human STS WI-11354, sequ
gb_pr:HSV351F8	+	196.00	312.66	3.9e-09	45678	Z70719 Human DNA sequence fr
gb_pr:BC003190	+	190.00	337.75	1.5e-10	793	BC003190 Homo sapiens, p75NTR-a
gb_pr:HUMOGC	+	190.00	336.71	1.8e-10	891	M38188 Human unknown protein fr
gb_pr:AF187064	+	190.00	336.71	1.8e-10	891	AF187064 Homo sapiens p75NTR-as
gb_pr:AK000959	+	187.00	328.38	5.2e-10	1229	AK000959 Homo sapiens CDNA FL
gb_pat:AX100231	+	187.00	327.45	5.8e-10	1364	AX100231 Sequence 32 from Pat
gb_pr:AF187066	+	187.00	292.48	5.1e-08	69648	AL035494 Human DNA sequence f
gb_to:AF187066	+	183.50	327.00	6.2e-10	700	AF187066 Mus musculus p75NTR-a
gb_to:AF097440	+	183.50	325.23	7.7e-10	854	AF097440 Mus musculus brain exp
gb_scs:G35294	-	170.00	305.77	9.4e-09	477	G35294 human STS SHGC-37409, se
gb_to:AF187065	+	150.50	289.43	9.9e-07	519	AF187065 Rattus norvegicus p75N
gb_scs:G72708	-	140.50	253.04	8.1e-06	421	G72708 MARC 4953-4954:991939031
gb_scs:G23964	-	103.50	186.62	0.0406	372	G23964 human STS WI-15922, sequ
gb_cm:OATRICH	+	93.50	139.71	16.63	9344	Z18361 O.aries trichohyalin.2
gb_ba:SCH69	+	93.50	127.76	77.00	35824	AL079308 Streptomyces coelic
gb_pr:AB040971	+	92.50	144.57	8.92	4408	AB040971 Homo sapiens mRNA for
gb_hcg:AC086858	+	92.50	119.94	210.09	70358	AC086858 Homo sapiens clone R
gb_hcg:AP003575	+	90.50	109.80	771.39	146017	AP003575 Oriza sativa chrom
gb_cm:OCTRICHA	+	90.00	133.54	36.71	9121	Z19092 O.cuniculus trichohyal
gb_ba:AF010496	-	88.00	102.92	1.9e+03	189370	AF010496 Rhodobacter capsula
gb_to:AB041591	-	87.50	144.14	9.43	1657	AB041591 Mus musculus brain c
gb_pat:AX188346	+	87.00	145.40	8.02	1298	AX188346 Sequence 4041 from Pa
gb_pat:A64152	+	87.00	141.54	13.16	2004	A64152 Sequence 18 from Patent
gb_pat:A68338	+	87.00	141.54	13.16	2004	A68338 Sequence 18 from Patent
gb_pat:AR031270	+	87.00	141.54	13.16	2004	AR031270 Sequence 18 from pat

gb_pat:AR036392	+	87.00	141.54	13.16	2004	!	AR036392	Sequence 18 from p									
gb_pat:AR060824	+	87.00	141.54	13.16	2004	!	AR060824	Sequence 18 from p									
gb_pat:AR065979	+	87.00	141.54	13.16	2004	!	AR065979	Sequence 18 from p									
gb_pat:AR088933	+	87.00	141.54	13.16	2004	!	AR088933	Sequence 18 from p									
gb_pat:A64169	+	87.00	139.30	17.52	2576	!	A64169	Sequence 35 from Pat									
seq_name: gb_pr:AF251053																	
seq_documentation_block:																	
LOCUS	AF251053	744 bp	mRNA		PRI		15-APR-2001										
DEFINITION	Homo sapiens X-linked protein mRNA, complete cds.																
ACCESSION	AF251053																
VERSION	AF251053.1	GI:13625167															
KEYWORDS																	
SOURCE	human.																
ORGANISM	Homo sapiens																
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.																
AUTHORS	1 (bases 1 to 744) Mao,Y., Xie,Y., Zhou,Z., Zhao,W., Wang,W., Huang,Y., Wang,S., Tang,R., Chen,X. and Wu,C.																
TITLE	Direct Submission																
JOURNAL	Submitted (29-MAR-2000) Institute of Genetics, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R.																
FEATURES	Location/Qualifiers																
source	1..744																
	/organism="Homo sapiens"																
	/db_xref="taxon:9606"																
	101..487																
	/codon_start=1																
	/product="X-linked protein"																
	/protein_id="AAK34943.1"																
	/db_xref="GI:13625168"																
	/translation="MESKEERALNLIVENVQENDEKEQVANKGEPLALPLNVS EYCVGRNRRFRVQPIYRWDMIRLGEPOARMRENERIGEEVRLMEKLREK QLSHSLRAVSTDPHHDHDFCLMP"																
BASE COUNT	222 a	134 c	208 g	180 t													
ORIGIN																	
alignment_scores:																	
	Quality:	679.00	Length:	128													
	Ratio:	5.305	Gaps:	0													
Percent Similarity:	100.000	Percent Identity:	100.000														
alignment_block:																	
US-09-327-750D-31 x AF251053																	
Align seg 1/1 to: AF251053 from: 1 to: 744																	
1	MetGluSerIysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa	17															
101	ATGGAGTCCAAAGAGAAAGACGGCTTAAACAATCTCATCTGGAAATGT	150															
17	laSnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG	34															
151	CAACACGAGAAATGATGAAAAAGATCAAAAGGACCAAGTTGCTAATAAG	200															
34	lyGluProLeuAlaLeuProLeuAsnValserGluTyrCysValProArg	50															
201	GGGAGCCCTTGGCCCTACCTTGAATGTTAGTGAATCTGTGTGCCTAGA	250															
51	GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr	67															
251	GGAAACCGTAGGCGGTTCCGCGTTAGGACGCCCATCTGCAGTATAGATG	300															
67	pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGlu	84															
301	GGACATAATGCATAGGCTTGGAGAGCCAGGCAAGGATGAGAGAGAGA	350															
84	snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg	100															

351 ATATGGAAGAGATTGGGAGAGAGTGAGACAGCTGTATGGAAGAGCTGAGG 400  
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117  
|||||  
401 GAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCACTGATCCCCCTCA 450  
|||||  
117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||  
451 CCATGACCATCAGCATGAGTTTGGCTTATGCCCC 484  
|||||

seq\_name: gb\_pat:AX078272

seq\_documentation\_block:  
LOCUS AX078272 898 bp DNA PAT 22-FEB-2001  
DEFINITION Sequence 76 from Patent WO0107471.  
ACCESSION AX078272  
VERSION AX078272.1 GI:13157963  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 898)  
AUTHORS Hillman,J.L., Lai,P., Tang,Y.T., Yue,H., Au-Young,J., Bandman,O.,  
TITLE Azimzai,Y., Yang,J., Lu,D.A., Baughn,M.R., Patterson,C. and Shah,P.  
JOURNAL Cell cycle and proliferation proteins  
PATENT: WO 0107471-A 76 01-FEB-2001;  
INCYTE Genomics, Inc. (US)  
FEATURES  
source  
1. .898  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 566415ACB1"

BASE COUNT 250 a 186 c 251 g 211 t  
ORIGIN  
alignment\_scores:  
Quality: 679.00 Length: 128  
Ratio: 5.305 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-327-750D-31 x AX078272 ..  
Align seg 1/1 to: AX078272 from: 1 to: 898

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17  
|||||  
254 ATGGAGTCCAAAGAGGAACGAGCGTTAAACAATCTCATCGTGGAAAAATGT 303  
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysC 34  
|||||  
304 CAACACGAGAAATGATGAAAGAGATGAAAGGAGCAAGTTGCTTAATAAAG 353  
34 lyGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
|||||  
354 GGGAGCCCTGGGCCCTACCTTGAATGTTAGTGAATCTGTGTGCTCTAGA 403  
51 GlyAsnArgArgArgPheArgValArgGlnProIleLeuGluTyrArgTr 67  
|||||  
404 GGAACCGTAGGCGGTTCGCGCTTAGCGACGCCCATCTCGCAGTATAGATG 453  
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84  
|||||  
454 GGACATAAGTCATAGGCTTGGAGAGCCACAGGCAAGGATGAGAGAGAGA 503  
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100  
|||||  
504 ATATGCAAGGATTTGGGAGGAGTGCAGACAGCTGATGGAAGAGCTGAGG 553  
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117  
|||||

554 GAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCACTGATCCCCCTCA 603  
117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||  
604 CCATGACCATCAGCATGAGTTTGGCTTATGCCCC 637  
|||||

seq\_name: gb\_pr:HSV870H8

seq\_documentation\_block:  
LOCUS HSV870H8 31321 bp DNA PRI 23-NOV-1999  
DEFINITION Human DNA sequence from cosmid V870H8, between markers DXS366 and  
DXS87 on chromosome X contains ESTs.

ACCESSION Z70233  
VERSION Z70233.1 GI:1235542  
KEYWORDS X.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 31321)  
AUTHORS Whiteley,M.  
JOURNAL Direct Submission  
COMMENT Submitted (19-MAR-1995) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk  
IMPORTANT: This sequence is not the entire insert of clone V870H8.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone V870H8 is at 1 in this sequence. The  
true left end of clone V693A8 is at 31221.  
V870H8 is from the human chromosome X-specific cosmid library.

FEATURES  
Location/Qualifiers  
1..31321  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="X"  
/clone="GHC-870H8"  
/clone\_lib="SCCV"  
475..701  
/note="L1 element fragment"  
802..988  
/note="L1 element fragment"  
2533..4138  
/note="L1 element fragment"  
/note="match: multiple ESTs"  
4189..4246  
/note="29 copies of 2 mer 91 % conserved"  
7496..7874  
/note="L1 element fragment"  
8022..8108  
/note="L1 element fragment"  
8302..8493  
/note="L1 element fragment"  
8585..8671  
/note="L1 element fragment"  
8769..9010  
/note="L1 element fragment"  
9259..9384  
/note="L1 element fragment"  
9624..9731  
/note="MLTIC element fragment"  
9781..9948  
/note="MLTIC element fragment"  
9890..9948  
/note="MLTIB element fragment"  
10418..10592  
/partial  
/note="Alu repeat: matches 298..108 of consensus"  
10593..10688  
/partial  
/note="Alu repeat: matches 96..1 of consensus"  
10698..10781  
/note="L1 element fragment"

```
repeat_region 11112..11403
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 11922..12004
/note="L1 element fragment"
repeat_region 12062..12308
/note="L1 element fragment"
repeat_region 12423..12707
/partial
/note="Alu repeat: matches 302..1 of consensus"
repeat_region 12710..12916
/note="L1 element fragment"
repeat_region 12979..13081
/note="L1 element fragment"
repeat_region 13473..13532
/note="MLT2A1 element fragment"
repeat_region 15871..15928
/note="L1 element fragment"
repeat_region 18180..18417
/note="L1 element fragment"
repeat_region 18778..19056
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 22279..22350
/note="L1 element fragment"
repeat_region 22449..22995
/note="L1 element fragment"
repeat_region 23239..23374
/note="L1 element fragment"
repeat_region 23371..23606
/note="MLT1B element fragment"
repeat_region 23395..23541
/note="MLT1A element fragment"
repeat_region 23705..23778
/note="MLT1A element fragment"
repeat_region 23805..23960
/note="L1 element fragment"
repeat_region 24011..24541
/note="L1 element fragment"
repeat_region 24638..24770
/partial
/note="Alu repeat: matches 150..1 of consensus"
repeat_region 24829..30975
/note="L1 element fragment"
BASE COUNT 9462 a 6147 c 6129 g 9583 t
ORIGIN

alignment_scores:
  Quality: 679.00      Length: 128
  Ratio: 5.305        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-327-750d-31 x HSV870H8

Align seg 1/1 to: HSV870H8 from: 1 to: 31321

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLleValGluAsnVa 17
3508 ATGAGTCCAAAGAGGAGGAGGTTAAACAATCTCATCGTGGAATGT 3557
|||||
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
3558 CAACACGAAATGATGAAAAAGATGAAAGAGGAGGAGTGTCTATTAAG 3607
|||||
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
3608 GGGAGCCCTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCTAGA 3657
|||||
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
3658 GGAACCGTAGGCGGTTCGCCGTAGGACCCCATCTCGCAGTATAGATG 3707
|||||
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
```

```
|||||
3708 GGACATAATGATAGCTTGGAGAGCCACAGGCAAGGATCAGAGAGAGA 3757
|||||
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
3758 ATATGAAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAAGCTGAGG 3807
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
3808 GAAACAGCTTGAGTCATAGTCTGGGGCAGTCAGCACTGATCCCCCTCA 3857
|||||
117 shiAspHisHisAspGluPheCysLeuMetPro 128
|||||
3858 CCATGACCATCAGCATGAGTTTGCCTTATGCC 3891

seq_name: gb_pr:AL133348

seq_documentation_block:
LOCUS AL133348 40584 bp DNA PRI 06-MAR-2000
DEFINITION Human DNA sequence from clone RPI-79p11 on chromosome Xq21.32-22.1.
Contains the gene for a novel protein similar to mouse Bex2
(brain-expressed X-linked protein 2), ESTs, STSs, GSSs and a
putative CpG island, complete sequence.
ACCESSION AL133348
VERSION AL133348.8 GI:7076386
KEYWORDS HTG; Bex2; CpG island.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 40584)
AUTHORS Wilson,S.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 24, 2000 this sequence version replaced gi:6997869.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RPI-79p11 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RPI-79p11 It may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap
between neighbouring submissions.
The true left end of clone LLOXNC01-105G4 is at 40485 in this
sequence. The true right end of clone LLOXNC01-177E8 is at 100 in
this sequence.
Location/Qualifiers
1..40584
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source
```



alignment\_scores:

Quality: 679.00 Length: 128  
Ratio: 5.305 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-31 x AL133348/rev ..

Align seg 1/1 to reverse of: AL133348 from: 1 to: 40584

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17  
|||||  
31132 ATGGATGCCAAAGAGGAGGAGCGTTAAACAATCTCATCTGGAATGT 31083  
|||||  
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34  
|||||  
31082 CAACGAGGAAATGATGAAAGAGATGAAAGGAGCAAGTTGCTAATAAG 31033  
|||||  
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
|||||  
31032 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCTAG 30983  
|||||  
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67  
|||||  
30982 GGAACCGTAGGCGGTTCGCGTTAGCGACCCATCTGCAGTATAGATG 30933  
|||||  
67 pAspIleMetHisArgLysGluGluProGlnAlaArgMetArgGluAla 84  
|||||  
30932 GGACATATGCTAGCTGGAGAGCCACAGGCAAGGATGAGAGAGGAGA 30883  
|||||  
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100  
|||||  
30882 ATATGGAAGGATTTGGGAGGAGTGCAGACAGCTGATGGAAGACTGAGG 30833  
|||||  
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117  
|||||  
30832 GAAAGACGTTGAGTCATAGTTTGGCGGCAGTCAGCACTGATCCCCCTCA 30783  
|||||  
117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||  
30782 CCATGACCATCAGATGAGTTTGGCTTATGCCCC 30749  
|||||

seq\_name: gb\_htg:HSU80B1

seq\_documentation\_block:

LOCUS HSU80B1 41029 bp DNA HTG 10-JUL-2001  
DEFINITION Homo sapiens chromosome X clone LLOXNC01-80B1 map q22.1-22.3, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.

ACCESSION

AL022169

VERSION

AL022169.3 GI:13276704

KEYWORDS

HTG; HTGS\_PHASE2; HTGS\_CANCELLED.

SOURCE

human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 41029)

REFERENCE

1 (bases 1 to 41029)

AUTHORS

Bird.C.

TITLE

Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,

JOURNAL

CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

request: clonerequest@sanger.ac.uk

COMMENT

On Mar 12, 2001 this sequence version replaced gi:12750927.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: cu80B1  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator ABI; 19% of reads

Chemistry: Dye-terminator Big Dye; 80% of reads  
Consensus quality: 41009 bases at least Q40  
Consensus quality: 41023 bases at least Q30  
Consensus quality: 41027 bases at least Q20  
Insert size: 41029; sum-of-contigs  
Insert size: 51751; 0.6% error; agarose-fp  
Quality coverage: 9.67x in Q20 bases; sum-of-contigs Quality  
coverage: 7.67x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES

Location/Qualifiers

1..41029  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="q22.1-22.3"  
/clone="LLOXNC01-80B1"  
/clone\_lib="LLOXNC01"

misc\_feature

1..41029  
/note="assembly\_fragment:00890"  
BASE COUNT 13249 a 7796 c 7781 g 12203 t  
ORIGIN

alignment\_scores:

Quality: 679.00 Length: 128  
Ratio: 5.305 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-31 x HSU80B1

Align seg 1/1 to: HSU80B1 from: 1 to: 41029

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17  
|||||  
18412 ATGGATGCCAAAGAGGAGGAGCGTTAAACAATCTCATCTGGAATGT 18461  
|||||  
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34  
|||||  
18462 CAACGAGGAAATGATGAAAGAGATGAAAGGAGCAAGTTGCTAATAAG 18511  
|||||  
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50  
|||||  
18512 GGGAGCCCTTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCTAGA 18561  
|||||  
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67  
|||||  
18562 GGAACCGTAGGCGGTTCGCGTTAGCGACCCATCTGCAGTATAGATG 18611  
|||||  
67 pAspIleMetHisArgLeuGluGluProGlnAlaArgMetArgGluAla 84  
|||||  
18612 GGACATATGCTAGCTGGAGAGCCACAGGCAAGGATGAGAGAGGAGA 18661  
|||||  
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100  
|||||  
18662 ATATGGAAGGATTTGGGAGGAGGTGAGACAGCTGATGGAAGACTGAGG 18711  
|||||  
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117  
|||||  
18712 GAAAGCAGTTGAGTCATAGTCTCGGGGAGTCAGCACTGATCCCCCTCA 18761  
|||||  
117 sHisAspHisHisAspGluPheCysLeuMetPro 128  
|||||  
18762 CCATGACCATCAGATGAGTTTGGCTTATGCCCC 18795  
|||||

seq\_name: gb\_pr:AF183416

seq\_documentation\_block:

LOCUS AF183416 642 bp mRNA PRI 02-SEP-2000

```

DEFINITION Homo sapiens ovarian granulosa cell 13.0 kDa protein hGR74 homolog
              mRNA, complete cds.
ACCESSION   AF183416
VERSION     AF183416.1 GI:9963770
KEYWORDS    ..
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 642)
AUTHORS     Peng,Y., Qian,B., Tu,Y., Xu,S., Han,Z., Fu,G. and Chen,Z.
TITLE       A novel gene expressed in human adrenal gland
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 642)
AUTHORS     Peng,Y., Gu,W., Huang,C., Xu,S., Han,Z., Fu,G. and Chen,Z.
TITLE       Direct Submission
JOURNAL     Submitted (03-SEP-1999) Chinese National Human Genome Center at
             Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
             Shanghai 201203, P. R. China
FEATURES             Location/Qualifiers
     source          1..642
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="adrenal gland"
     CDS              3..380
                     /note="hGR74-h protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="ovarian granulosa cell 13.0 kDa protein hGR74
                     homolog"
                     /protein_id="AA093685.1"
                     /db_xref="GI:9963771"
                     /translation="MESKEKRAVNSLSMENANQNEEKEQVANKCEPLALPLDAGEYC
                     VPRGNRRFRVROPILQYRWDMHRLGEPQARMRENNMERIGEEVRLMEKLRKQLS
                     HSLRAVSTDPHHDHDFCLMP"
     BASE COUNT      204 a 118 c 157 g 163 t
     ORIGIN
101  MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
111  |||||
121  |||||
131  |||||
141  |||||
151  |||||
161  |||||
171  |||||
181  |||||
191  |||||
201  |||||
211  |||||
221  |||||
231  |||||
241  |||||
251  |||||
261  |||||
271  |||||
281  |||||
291  |||||
301  |||||
311  |||||
321  |||||
331  |||||
341  |||||
351  |||||
361  |||||
371  |||||
381  |||||
391  |||||
401  |||||
411  |||||
421  |||||
431  |||||
441  |||||
451  |||||
461  |||||
471  |||||
481  |||||
491  |||||
501  |||||
511  |||||
521  |||||
531  |||||
541  |||||
551  |||||
561  |||||
571  |||||
581  |||||
591  |||||
601  |||||
611  |||||
621  |||||
631  |||||
642  |||||
3  ATGGAGTCCAAAGAGAACGAGCAGTAAACAGTCTCAGCATGGAATGC 52
17  lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
53  CAACCAAGAAAAAT.....GAAGAAAAGGAGCAAGTTGCTTAATAAAG 93
34  lYGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
94  GGGAGCCCTTGGCCCTCCCTTTGGATGCTGCTGAATCTGTGTCCTAGA 143
51  GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
144  GGAATCGTAGCGGTTCGCGTTAGGCAGCCCATCTCGCAGTATAGATG 193
67  pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluAla 84
194  GGATATGATCATAGGCTTGAGAACACACAGGCAAGGATGAGAGAGA 243
84  snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
244  ATATGGAAGAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGAGG 293

```

```

101  GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProCohI 117
111  |||||
121  |||||
131  |||||
141  |||||
151  |||||
161  |||||
171  |||||
181  |||||
191  |||||
201  |||||
211  |||||
221  |||||
231  |||||
241  |||||
251  |||||
261  |||||
271  |||||
281  |||||
291  |||||
301  |||||
311  |||||
321  |||||
331  |||||
343  GAAAAGCAGTTGAGTCATAGCTCGCGGCGAGTCAGCACTGACCCCTCA 343
117  sHisAspHisHisAspGluPheCysLeuMetPro 128
134  CCATGACCATCATGATGAGTTTGGCTTATGCC 377
seq_name: gb_pr:AF220189
seq_documentation_block:
LOCUS       AF220189             828 bp    mRNA             PRI      04-MAY-2000
DEFINITION Homo sapiens uncharacterized hypothalamus protein HBEX2 mRNA,
complete cds.
ACCESSION   AF220189
VERSION     AF220189.1 GI:7689028
KEYWORDS    SOURCE
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 828)
AUTHORS     Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.
TITLE       A novel gene expressed in human hypothalamus
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 828)
AUTHORS     Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.
TITLE       Direct Submission
JOURNAL     Submitted (30-DEC-1999) Chinese National Human Genome Center at
             Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
             Shanghai 201203, P. R. China
FEATURES             Location/Qualifiers
     source          1..828
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="hypothalamus"
                     /protein_id="AF067654.1"
                     /db_xref="GI:7689029"
                     /translation="MESKEKRAVNSLSMENANQNEEKEQVANKCEPLALPLDAGEYC
                     VPRGNRRFRVROPILQYRWDMHRLGEPQARMRENNMERIGEEVRLMEKLRKQLS
                     HSLRAVSTDPHHDHDFCLMP"
     BASE COUNT      256 a 170 c 221 g 181 t
     ORIGIN
alignment_scores:
  Quality: 606.50          Length: 128
  Ratio: 4.891            Gaps: 1
  Percent Similarity: 96.875 Percent Identity: 89.062
alignment_block:
US-09-327-750D-31 x AF220189
Align seg 1/1 to: AF220189 from: 1 to: 828
1  MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
167  ATGGAGTCCAAAGAGAACGAGCAGTAAACAGTCTCAGCATGGAATGC 216
17  lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
217  CAACCAAGAAAAAT.....GAAGAAAAGGAGCAAGTTGCTTAATAAAG 257
34  lYGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
258  GGGAGCCCTTGGCCCTCCCTTTGGATGCTGCTGAATCTGTGTCCTAGA 307
51  GlyAsnArgArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
308  GGAATCGTAGCGGTTCGCGTTAGGCAGCCCATCTCGCAGTATAGATG 357

```





```
repeat_region 29603..30740
/note="TRIGGER1 repeat: matches 587..1757 of consensus"
repeat_region 31027..31365
/note="TRIGGER1 repeat: matches 1757..2105 of consensus"
repeat_region 31684..31891
/note="TRIGGER1 repeat: matches 2105..2132 of consensus"
repeat_region 31809..32015
/note="TRIGGER1 repeat: matches 2132..2246 of consensus"
repeat_region 32323..32389
/note="TRIGGER1 repeat: matches 2246..2313 of consensus"
repeat_region 34353..34636
/note="MLT1-INTERNAL repeat: matches 607..889 of consensus"
repeat_region 35149..35376
/note="MLT1-INTERNAL repeat: matches 1001..1233 of consensus"
repeat_region 35523..35670
/note="MLT1C repeat: matches 8..159 of consensus"
BASE COUNT 9708 a 7094 c 6805 g 12107 t
ORIGIN

alignment_scores:
  Quality: 606.50      Length: 128
  Ratio: 4.891         Gaps: 1
  Percent Similarity: 96.875 Percent Identity: 89.062

alignment_block:
US-09-327-750d-31 x HS198P4/rev ..
Align seg 1/1 to reverse of: HS198P4 from: 1 to: 35714
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
24358 ATGGAGTCCAAAGAGAAACAGCAGTAAACAGTCTCAGCATGGAATGC 24309
|||||
17 lAsnGlnGluAsnAspGluLysGluGlnValAlaAsnLysG 34
|||||
24308 CAACCAAGAAAT.....GAAGAAAGGAGCAAGTTGCTAATAAG 24268
|||||
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
24267 GGGAGCCCTTGGCCCTCCCTTGGATGCTGGTGAATACTGTGTGCCTAGA 24218
|||||
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
24217 GGAATCGTAGGCGGTTCGCGGTAGGCAGCCCATCTCGCAGTATAGATG 24168
|||||
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
|||||
24167 GGATATGATGATGCTTGGAGAACACAGGCAAGGATGAGAGAGAGA 24118
|||||
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
24117 ATATGAAAGGATGGGAGGAGGTGAGACAGCTGATGAAAGGCTGAGG 24068
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
24067 GAAAGACAGTTGAGTCATAGTCTGCGGGCAGTCAGCACTGACCCCTCA 24018
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
24017 CCATGACCATCATGATGAGTTTGCCTTATGCCC 23984
|||||

seq_name: gb_pr:AF237783
seq_documentation_block:
LOCUS AF237783 791 bp mRNA PRI 02-SEP-2000
DEFINITION Homo sapiens brain-expressed protein BEX1 (BEX1) mRNA, complete cds.
ACCESSION AF237783
VERSION AF237783.1 GI:9963898
KEYWORDS
```

```
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 791)
AUTHORS Yang,Q.S., Ying,K., Xie,Y. and Mao,Y.M.
TITLE A Novel Human X-linked Brain Expressed Protein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 791)
AUTHORS Mao,Y.M., Xie,Y., Yang,Q.S., Wu,H., Lin,S. and Ying,K.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2000) Genetic Research Institute, Fudan
University, 220 Handan Rd., Shanghai 200433, P.R.China
FEATURES
    Location/Qualifiers
        1..791
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="X"
            /map="between DXS990 and DX1059 by RH (Stanford G3)"
            1..791
            /gene="BEX1"
            172..549
            /gene="BEX1"
            /codon_start=1
            /product="brain-expressed protein BEX1"
            /protein_id="AAG09752.1"
            /db_xref="GI:9963899"
            /translation="MESKEKLVNSLSMENANQNEKEQVANKGEPLALPDAGEYC
            VPRGNRRFRVRQPIQYRMDMMHRLGEPQARMRENMERIGEGVRLMEXLREKQLS
            HSLRAVSTDPDHPHHDHDFCLMP"
        BASE COUNT 214 a 172 c 218 g 187 t
        ORIGIN

alignment_scores:
  Quality: 592.50      Length: 128
  Ratio: 4.857         Gaps: 1
  Percent Similarity: 95.312 Percent Identity: 87.500

alignment_block:
US-09-327-750D-31 x AF237783 ..
Align seg 1/1 to: AF237783 from: 1 to: 791
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
172 ATGGAGTCCAAAGAGAAACAGTAAACAGTCTCAGCATGGAATGC 221
|||||
17 lAsnGlnGluAsnAspGluLysGluGlnValAlaAsnLysG 34
|||||
222 CAACCAAGAAAT.....GAAGAAAGGAGCAAGTTGCTAATAAG 262
|||||
34 lylGluProLeuAlaLeuProLeuAsnValSerGluTyrCysValProArg 50
|||||
263 GGGAGCCCTTGGCCCTCCCTTGGATGCTGGTGAATACTGTGTGCCTAGA 312
|||||
51 GlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrArgTr 67
|||||
313 GGAATCGTAGGCGGTTCGCGGTAGGCAGCCCATCTCGCAGTATAGATG 362
|||||
67 pAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluGluA 84
|||||
363 GGATATGATGATGCTTGGAGAACACAGGCAAGGATGAGAGAGAGA 412
|||||
84 snMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeuArg 100
|||||
413 ATATGAAAGGATGGGAGGAGGTGAGACACCTGATGGAAGGCTGAGG 462
|||||
101 GluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
463 GAAAGACAGTTGAGTCATAGTCTGCGGGCAGTCAGCACTGACCCCTCA 512
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
```

```

|||||
513 CCATGACCATCATGATGAGTTTGCCTTATGCC 546

seq_name: gb_ro:AF097439

seq_documentation_block: 785 bp mRNA ROD 13-APR-1999
LOCUS AF097439 Mus musculus brain expressed x-linked protein 2 (Bex2) mRNA,
DEFINITION complete cds.
ACCESSION AF097439
VERSION AF097439.1 GI:4580591
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 785)
Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
JOURNAL is a member of a novel gene family on the mouse X chromosome
MEDLINE Hum. Mol. Genet. 8 (4), 611-619 (1999)
REMARK 99172070
Erratum:[published erratum appears in Hum Mol Genet 1999
May;8(5):943]
REFERENCE 2 (bases 1 to 785)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
source
1..785
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="dbEST:W48832"
/chromosome="X"
/map="near Plp"
/dev_stage="embryo; 15.5 dpc"
1..785
/gene="Bex2"
139..528
/gene="Bex2"
/codon_start=1
/product="brain expressed x-linked protein 2"
/protein_id="AAD24430.1"
/db_xref="GI:4580592"
/translation="MESKVEQGVKNLNMENDHQEKEEKPDQASKRDPVALPFEA
GDYVPRGRRRRFRVQPIVIRWDLMRVGPQGRMEENVQRFQDDVROLMEKLR
RQLSHSLRAVSTDPHHDHDECLMP"
BASE COUNT 212 a 163 c 235 g 175 t
ORIGIN

alignment_scores:
Quality: 502.50 Length: 129
Ratio: 4.153 Gaps: 1
Percent similarity: 93.798 Percent Identity: 71.318

alignment_block:
US-09-327-750d-31 x AF097439 ..
Align seg 1/1 to: AF097439 from: 1 to: 785
1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17
|||||
139 ATGGAGTCCAAATGGCAACAGCGGTGAAATAATCTCAACATGGAGATGA 188
|||||
17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34
|||||
189 CCATCAGGAAGAGGAGGAAGGAAGAACAGCCACAGGATGCTAGCAAAA 238
|||||
34 lyGluProLeu...AlaLeuProLeuAsnValSerGluTyrCysValPro 49
|||||

```

---

```

239 GGGATCCGATTGTGGCCCTGCCTTTCGAAGCTGGAGACTACTACGTGCCT 288
50 ArgGlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrAr 66
|||||
289 AGAGGAGCTCGCAGCGGTTCGGGTTCGCAGCCCATCGTGCCTACAG 338
66 gTTPAspIleMetHisArgLeuGluProGlnAlaArgMetArgGluG 83
|||||
339 ATGGGACCTGATCATAGGTTGGGAGGCCAGGAGGAGGATGAGAGG 388
83 luAsnMetGluArgIleGlyGluValArgGlnLeuMetGluLysLeu 99
|||||
389 AGAACGTACAGAGTTTGGGGATGATGTGACAGCTCATGAGAGAGCTG 438
100 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||
439 AGGAAAGCAGCTGAGCACAGCCTCGCGCGGTAGCACTGACCCGCC 488
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
489 TCATCATGACCACTGATGAGTTTGCCTTATGCC 525
seq_name: gb_ro:AF097437

seq_documentation_block:
LOCUS AF097437 2269 bp DNA ROD 05-AUG-1999
DEFINITION Mus musculus Bex1 protein (Bex1) gene, complete cds.
ACCESSION AF097437
VERSION AF097437.1 GI:5702155
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2269)
Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
JOURNAL is a member of a novel gene family on the mouse X chromosome.
MEDLINE Hum. Mol. Genet. 8 (4), 611-619 (1999)
REMARK 99172070
PUBMED 10072429
REFERENCE 2 (bases 1 to 2269)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4030, Australia
FEATURES
source
1..2269
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"
/chromosome="X"
/map="near Plp"
/tissue_type="liver"
Join(409..509,866..942,1263..1873)
/gene="Bex1"
/product="Bex1 protein"
409..1873
/gene="Bex1"
/note="expressed in brain; X-linked"
1268..1654
/gene="Bex1"
/codon_start=1
/product="Bex1 protein"
/protein_id="AAD47168.1"
/db_xref="GI:5702156"
/translation="MESKDGQVKLNLMENDHQEKEEKPDQTIIRREPVALTSEAG
KNCAPRGRRRRFRVQPIAHYRWDLMRVGPQGRMEENVQRFQDDVROLMEKLR
QLSHSLRAVSTDPHHDHDECLMP"
BASE COUNT 563 a 554 c 705 g 447 t
ORIGIN

```



TITLE Characterization of genes which exhibit reduced expression during the retinoic acid-induced differentiation of F9 teratocarcinoma cells: involvement of cyclin D3 in RA-mediated growth arrest

JOURNAL Mol. Cell. Endocrinol. 143 (1-2), 155-166 (1998)

MEDLINE 99021197

REFERENCE 2 (bases 1 to 838)

AUTHORS Faria,T.N., Larosa,G., Wilen,E., Liao,L. and Gudas,L.J.

TITLE Direct Submission

JOURNAL Submitted (26-FEB-1998) Pharmacology, Cornell University Medical College, 1300 York Avenue, New York, NY 10021, USA

FEATURES

source

1..838

/organism="Mus musculus"

/db\_xref="taxon:10090"

/cell\_line="F9 teratocarcinoma"

229..579

/note="retinoic acid reduced gene expression"

/codon\_start=1

/product="REX-3"

/protein\_id="AAC61929.1"

/db\_xref="GI:3510643"

/translation="MENDHOKKEEKEKPODTIRREPAVALISEAGKNCAPRGRRRRFVRQPIAHYRWDLQVRGEPQGRMRNEVQFGDVRQLMKLRQLSHSLRAVSTDPPHDDHDEFLMP"

BASE COUNT 238 a 181 c 251 g 168 t

ORIGIN

alignment\_scores

Quality: 450.00 Length: 129

Ratio: 3.879 Gaps: 2

Percent Similarity: 89.922 Percent Identity: 66.667

alignment\_block:

US-09-327-750D-31 x AF051347

Align seg 1/1 to: AF051347 from: 1 to: 838

1 MetGluSerLysGluGluArgAlaLeuAsnAsnLeuIleValGluAsnVa 17

193 ATGGAGTCCAAAGATCAA...GGCGTGAAAAATCTCAACATGGAGATGA 239

17 lAsnGlnGluAsnAspGluLysAspGluLysGluGlnValAlaAsnLysG 34

240 CCATCAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 289

34 lyGluPro...LeuAlaLeuProLeuAsnValSerGluTyrCysValPro 49

290 GGGAGCCAGCTGTGGCCCTGATCTCCGAGGCTGGCAAAACCTGTGGCCT 339

50 ArgGlyAsnArgArgPheArgValArgGlnProIleLeuGlnTyrAr 66

340 AGAGGAGTCCGAGCGGTTCGGGTTCGGCAGCCCATCTCCTCATATAG 389

66 gTrpAspIleMetHisArgLeuGlyGluProGlnAlaArgMetArgGluG 83

390 ATGGGACCTCATCAGAGGTTGGGAGGCCCGCCAGGAGGATCAGAGAGG 439

83 luAsnMetGluArgIleGlyGluGluValArgGlnLeuMetGluLysLeu 99

440 AGAACCTACAGAGGTTGGGAGGTTGATCTGAGACAGCTCATGGAGAGCTG 489

100 ArgGluLysGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116

490 AGGGAAGGAGGAGTGGAGCCACCGCTGGCGGCGGTAGCACTGACCCGCC 539

116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128

540 TCATCATGACCACCATCATGATGAGTTTGCCTCATGCC 576

seq\_name: gb\_ro:BC003254

seq\_documentation\_block:

LOCUS BC003254 2901 bp mRNA ROD 20-FEB-2001

DEFINITION Mus musculus, Similar to dentatorubral pallidolusyan atrophy, clone MGC:5758, mRNA, complete cds.

ACCESSION BC003254

VERSION BC003254.1 GI:13096906

KEYWORDS MGC.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2901)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (20-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: villalona@bcm.tmc.edu.  
Villalona, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 9 Row: 0 Column: 6.

FEATURES

Location/Qualifiers

1..2901

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="MGC:5758 IMAGE:3500522"

/tissue\_type="Mammary tumor. C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."

/clone.lib="NCI CGAP\_Mam6"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

755..2359

/codon\_start=1

/product="Similar to dentatorubral pallidolusyan atrophy"

/protein\_id="AAH03254.1"

/db\_xref="GI:13096907"

/translation="MVQFLLPTLPGLPKPPIHVHTPLHPRAPKEHTPPSHQSQSP PPQLPPLSPPAATTGPPLTATQIKQEPAEVEPEPPSPVPKSPSPKPPKVDVPS HASQSAFNKHLDRGNSCARSDLYFVPLEGSKLAKRADLVKVRREAEQRAREEKE REREREREKERERELERSVKLAQEGRAVCEPGLGVPVHRPFPFGSAVATVPP YLGPDTPALRTLSEYARPHVMSPGNRNHPFYVPLGAVDPLGLLYNVPALYSSDPAARE REREARERDLRDLKPGFEVKSPELEPLHGVPGCLDPEPRHGGALQPGPGLHPFP FHPSLGPLERLERLAAQAPALPMSYAEERLAERQAERVAALGNDRLARLQMLNVT PPHQHSHSHSLHLHQDDATHAASASVHPLDPLASGSHLTRIDPYTPAGTLPNPLPH PLHENVLRHQLFAAPYRDLPAISAPMSAAHQIAQMAHQSAELQRLALEQQQWLHAH HPLHSVLPQAQEDYYSHLKESDKPL"

BASE COUNT 623 a 958 c 791 g 529 t

ORIGIN

alignment\_scores

Quality: 388.00 Length: 119

Ratio: 3.660 Gaps: 2

Percent Similarity: 89.076 Percent Identity: 63.866

alignment\_block:

US-09-327-750D-31 x BC003254

Protocol:   
 Template: 10 ng

47	CysValProArgGly	AsnArgArgArgPhe	ArgValArgGlnProIle	62
503	TGTCCCTTANAGAAACC	CTGTAGCGGTCTCCCGCTTAGGCANCC	CCAC	454
63	LeuGlnTyrArgTrpAspIle	MetHisArgLeuGlyGlu	ProGlnAla	79
453	CCGCAGTANAGATGGGATATGATGCATYAGGCTT	CGAGAACCCACAGGCCAA	404	
79	rgMetArgGluGluAsnMetGlu	ArgIleGlyGluGluValArgGlnLeu	95	
403	GGATCAGACAGAGAGAAATATGAAAGGATATGGGGAGGAGGT	TCAGACAGCTG	354	
96	MetGluLysLeuArgGluLysGlnLeuSerHisSerLeuArgAlaVal	112		
353	ATGGAAAGCTGAGGGGAAACGACGTTGAGTCANAGTCTCGCGGCGTCAG	304		
112	rThrAspProProHisHisAspHisHisAspGluPheCysLeuMetPro	128		
303	CACCTACCCCTCACCATGACCATCATGATGAGCTTTGCTTTATGCC	255		



```

34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
242 AGCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACTGTGGCCCTAGA 291
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
292 GGAGGTGCGAGCGGTTCGGGGTTCGGCAGCCCATCGCTCACTATAGTG 341
|||||
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGlu 84
|||||
342 GGACCTGATCGACAGGCTGGGAGGCCCCAGGGAAGGATGAGAGAGAGA 391
|||||
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||
392 ACGTACAGAGTTTGGGGTGTGTGAGACAGCTCATGGAGAAGCTGAGG 441
|||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
442 GAAAGGCAGCTGAGCCACAGCCTCGCGGGGTGTAGCACTGACCCGCTCA 491
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
492 TCATGACCCACCATGATGAGTTTGGCTCATGCCCC 525
|||||

seq_name: gb_est1:AW536404

seq_documentation_block:
LOCUS AW536404 642 bp mRNA EST 31-AUG-2000
DEFINITION G0104A11-3 NIA Mouse E7.5 Embryonic Portion cDNA Library Mus
musculus cDNA clone G0104A11 3', mRNA sequence.
ACCESSION AW536404
VERSION AW536404.1 GI:7178821
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac
M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Dol, H., Wood, W.H.,
III, Becker, K.G. and Ko, M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
20381348
CONTACT: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: G0104 row: A column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 642
POLYA=Yes.

FEATURES
source
1..642
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G0104A11"
/clone_lib="NIA Mouse E7.5 Embryonic Portion cDNA Library"
/sex="unknown"
/dev_stage="7.5dpc Embryo"
/lab_host="DH10B"
/notes="vector: pSPORT1 (Gibco/BRL Life Technology);
Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
6 Embryo. The double-stranded cDNA was synthesized by
Gibco's kit with an Oligo(dT) primer (NotI primer-adaptor
from GibcoBRL)
[5'-pGACTAGTCTTAGATCCGAGCGGCCCTTTT-3']
from 0.5ug of mRNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by

```

ethanol-precipitation. The cDNAs were ligated to Lone-linker L1-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang and Minoru S. H. Ko. "

BASE COUNT 145 a 182 c 133 g 182 t  
ORIGIN

alignment\_scores:  
Quality: 683.00 Length: 128  
Ratio: 5.378 Gaps: 0  
Percent Similarity: 99.219 Percent Identity: 99.219

alignment\_block:  
US-09-327-750D-30 x AW536404/rev ..

Align seg 1/1 to reverse of: AW536404 from: 1 to: 642

```

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH 17
|||||
606 ATGGAGTCCAAAGATCAAGGCGTGAATAATCTCAACATGGAGAAATGACCA 557
|||||
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArgG 34
|||||
556 TCAGAAAAGAGGAGAGAGAAAACCAAGATACCATCAGAGGG 507
|||||
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
506 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 457
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
456 GGAGGTGCGAGCGGTTCGGGGTTCGGCAGCCCATCGCTCACTATAGTG 407
|||||
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGlu 84
|||||
406 GGACCTGATGACAGAGGTTGGGAGGCCCGGAGGATGAGAGAGAGA 357
|||||
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||
356 ACGTACAGAGTTTGGGGGTGTGTGAGACAGCTCATGGAGAAGCTGAGG 307
|||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
306 GAAAGGCAGCTGAGCCACAGCCTCGCGGGGTGTAGCACTGACCCGCTCA 257
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
256 TCATGACCCACCATGATGAGTTTGGCTCATGCCCC 223
|||||

seq_name: gb_est2:BG064920

seq_documentation_block:
LOCUS BG064920 707 bp mRNA EST 26-JAN-2001
DEFINITION H3025D01-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3025D01 3', mRNA sequence.
ACCESSION BG064920
VERSION BG064920.1 GI:12547483
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 707)

```



## AUTHORS

Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka  
,T.S., Carter,M.G. and KO,M.S.H.

## TITLE

Verification and Initial annotation of NIA mouse 15K cDNA clone set  
Unpublished (2001)

## JOURNAL

Other ESTs: H3025D01-5

## COMMENT

Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please

visit <http://lsun.grc.nia.nih.gov/cDNA/15k.html> for details.

Plate: H3025 row: D column: 01

Seq primer: -21M13 Forward

High quality sequence stop: 707

POLYA=Yes.

## FEATURES

Location/Qualifiers

1..707

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="H3025D01"

/clone\_lib="NIA Mouse 15K cDNA Clone Set"

/sex="Clones arrayed from a variety of cDNA libraries"

/dev\_stage="Clones arrayed from a variety of cDNA

libraries"

/lab\_host="DH10B"

/note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; This

clone is among a rearranged set of 15,247 clones from 11

embryo cDNA libraries (including preimplantation stage

embryos from unfertilized egg to blastocyst, embryonic

part of E7.5 embryos, extraembryonic part of E7.5 embryos

, and E12.5 female mesonephros/gonad) and one newborn

ovary cDNA library. Average insert size 1.5 kb. All

source libraries are cloned unidirectionally with Oligo(dT

)-Not primers. References include: (1) Genome-wide

expression profiling of mid-gestation placenta and embryo

using a 15,000 mouse developmental cDNA microarray, 2000,

Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)

Large-scale cDNA analysis reveals phased gene expression

patterns during preimplantation mouse development, 2000,

Development, 127: 1737-1749; (3) Genome-wide mapping of

unselected transcripts from extraembryonic tissue of

7.5-day mouse embryos reveals enrichment in the t-complex

and under-representation on the X chromosome, 1998, Hum

Mol Genet 7: 1967-1978."

Mol Genet 7: 1967-1978"

154 a 209 c 146 g 198 t

BASE COUNT

ORIGIN

## alignment\_scores:

Quality: 680.00 Length: 128

Ratio: 5.354 Gaps: 0

Percent Similarity: 99.219 Percent Identity: 98.438

## alignment\_block:

US-09-327-750d-30 x BG064920/rev ..

Align seg 1/1 to reverse of: BG064920 from: 1 to: 707

1 luMetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHi 17

|||||

606 ATGGAGTCCAAAGATCAAGCGCTGAAATCTCAACATGGAGAAATGACCA 557

|||||

17 sGlnLysLysGluLysGluLysGluLysProGlnAspThrIleArgArgG 34

|||||

556 TCAGAAAAAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 507

|||||

34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50

|||||

506 ACCGAGCTGTGCCCTGACCTCCGAGGCTGGCAAAACTGTGCACCTAGA 457

|||||

51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67

|||||

456 GGAGGTGCGAGCGGTTCGGGTTCGGCAGCCCATCGCTCACTATAGATG 407

|||||

67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGluA 84

|||||

406 GGACCTGATGCAGAGGTTGGGGAGCCCGGAGGATGAGAGAGAGA 357

|||||

84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100

|||||

356 ACGTACAGAGGTTTGGGGTGTATGTGAGACAGCTCATGAGAGAGCTGAGG 307

|||||

101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117

|||||

306 GAAAGGACAGCTGAGCCACAGCGTCGGGCGGTAGCACTGACCCGCTCA 257

|||||

117 shiAspHisAspGluPheCysLeuMetPro 128

|||||

256 TCATGACCACCATGATGAGTTTTCCTCATGCC 223

|||||

seq\_name: gb\_est2:BG277659

seq\_documentation\_block:

LOCUS BG277659 575 bp mRNA EST 21-FEB-2001

DEFINITION ux47c11.y1 Soares\_NMAX\_maxillary\_process Mus musculus cDNA clone

IMAGE:3513237 5' similar to TR:09R1J2 Q9R1J2 BEX1 PROTEIN. ;, mRNA

sequence.

ACCESSION BG277659

VERSION BG277659.1 GI:13073183

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 575)

REFERENCE NC1-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Other ESTs: ux47c11.x1

Contact: Robert Strausberg, Ph.D.

Email: cgaos-r@mail.nih.gov

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

MGI:1393813

Seq primer: -40Rp from Gibco

High quality sequence stop: 462.

Location/Qualifiers

1..575

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:3513237"

/clone\_lib="Soares\_NMAX\_maxillary\_process"

/tissue\_type="maxillary process"

/lab\_host="DH10B (phage-resistant)"

/note="vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site.1: NotI; Site.2: EcoRI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M. Fatima Bonaldo. "

156 a 135 c 201 g 81 t 1 others

BASE COUNT

ORIGIN

## alignment\_scores:

Quality: 670.00 Length: 128

Ratio: 5.317 Gaps: 0

Percent Similarity: 98.438 Percent Identity: 97.656

## alignment\_block:

US-09-327-750D-30 x BG277659

Align seg 1/1 to: BG277659 from: 1 to: 575

```
1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17
|||||
192 ATGGAGTCCAAAGATCAAGCGCTGAAATAATCTCAACATGGAGAATGACCA 241
|||||
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArgG 34
|||||
242 TCAGAAAAGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 291
|||||
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
292 AGCCAGCTGTGCGCCCTGACCTCCGAGGCTGGCAAAATCTGCACCTAGA 341
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
342 GGAGGTCCGAGGCGGTTCGGGGTTCGGCAGCCCATCGCTCACTATAGATG 391
|||||
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluAla 84
|||||
392 GGACCTGTGTCAGAGGTTGGGAGGCTGGGAGGAGGAGGAGGAGGAGG 441
|||||
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||
442 ACCTACAGAGTTTNGGGTGTGATGTGAGACAGCTCATGGAGAGCTGAGG 491
|||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
492 GAAAGCAGCTGAGCCACAGCTCGGGGGTTCGACCTGAGCCCGCTCA 541
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
542 TCATGACCAACCATGATGAGTTGTGCTCATGCC 575
|||||
```

seq\_name: gb\_est2:BG870503

```
seq_documentation_block: 756 bp mRNA EST 29-MAY-2001
LOCUS BG870503 602791422f1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4922644 5',
DEFINITION mRNA sequence.
ACCESSION BG870503
VERSION BG870503.1 GI:14221043
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 756)
NTH-MSC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10842 row: 1 column: 05
High quality sequence stop: 756.
Location/Qualifiers
1..756
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
```

## FEATURES

source

NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 207 a 164 c 230 g 155 t  
ORIGIN

alignment\_scores:  
Quality: 670.00 Length: 129  
Ratio: 5.276 Gaps: 1  
Percent Similarity: 98.450 Percent Identity: 98.450

alignment\_block:

US-09-327-750D-30 x BG870503

Align seg 1/1 to: BG870503 from: 1 to: 756

```
1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17
|||||
165 ATGGAGTCCAAAGATCAAGCGCTGAAATAATCTCAACATGGAGAATGACCA 214
|||||
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArgG 34
|||||
215 TCAGAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 264
|||||
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
265 ACCCAGCTGTGCGCCCTGACCTCCGAGGCTGGCAAAATCTGCACCTAGA 314
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
315 GCAGGTCCGAGGCGGTTCGGGGTTCGGCAGCCCATCGCTCACTATAGATG 364
|||||
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMet.ArgGluGlu 83
|||||
365 GGACCTGTGTCAGAGGTTGGGAGGCTGGGAGGAGGAGGAGGAGGAGG 414
|||||
84 AsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuAr 100
|||||
415 AACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTCATGGAGAGCTGAG 464
|||||
100 gGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProH 117
|||||
465 GGAAGGCGAGCTGAGCCACAGCTCGGGCGGTTCGACCTGAGCCCGCTC 514
|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
515 ATCATGACCAACCATGATGAGTTTGTGCTCATGCC 549
|||||
```

seq\_name: gb\_est1:BE654459

```
seq_documentation_block: 499 bp mRNA EST 06-SEP-2000
LOCUS BE654459 499 bp mRNA EST 06-SEP-2000
DEFINITION UT-M-AJ1-aha-f-10-0-UI.r1 NIH_BMAP_MOB_N Mus musculus cDNA clone
UT-M-AJ1-aha-f-10-0-UI 5', mRNA sequence.
ACCESSION BE654459
VERSION BE654459.1 GI:9980372
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 499)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
JOURNAL 9704477  
MEDLINE Contact: Chin, H  
COMMENT National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706

Fax: 301 443 9890

Email: mEst@mail.nih.gov

cDNA Library Preparation: M.B. Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Reverse.

Seq primer: M13 Reverse.

# FEATURES:

source

Location/Qualifiers

1..499

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UI-M-AJ1-aha-f-10-0-UI"

/clone\_lib="NIH\_BMAP\_MOB\_N"

/dev\_stage="27-32 days"

/lab\_host="DHI08 (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The

NIH\_BMAP\_MOB\_N library is a normalized library constructed

from mouse olfactory bulbs. The tag is a string of 5

nucleotides present between the Not I site and the

oligo-dr track. The library was constructed as described

by Bonaldo, Lennon and Soares, Genome Research 6: 791-806

, 1996. Tissue provided by Ms. Annie Novakovich,

Zivic-Miller Laboratories."

146 a 108 c 179 g 56 t

BASE COUNT

ORIGIN

alignment\_scores:

Quality: 652.00

Ratio: 5.344

Percent Similarity: 99.187

Percent Identity: 99.187

alignment\_block:

US-09-327-750D-30 x BE654459

Align seg 1/1 to: BE654459 from: 1 to: 499

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHi 17

|||||

131 ATGGAGTCCAAAGATCAAGCGCTGAAAAATCTCAACATGGAGAAATGACCA 180

|||||

17 scGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArg 34

|||||

181 TCAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 230

|||||

34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50

|||||

231 ACCCAGCTGTGCGCTGACCTCCGAGGCTGGCAGCCCAAACTGTGCACCTAGA 280

|||||

51 GlyGlyArgArgArgPheArgValArgGlnProIleAlaHisTyArgTr 67

|||||

281 GGAGTGCAGCGGTTCGGGTTCGGGTTCGGGTTCGGGTTCGGGTTCGGGTTC 330

|||||

67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGlu 84

|||||

331 GGACCTGATCGACAGGTTGGGAGCCCGCCAGGAGGATGAGAGGAGGAGA 380

|||||

84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100

|||||

381 ACCTACAGAGGTTGGGGGTGATGTGAGACAGCTCATGGAGAAGCTGAGG 430

|||||

101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117

|||||

431 GAAAGCAGCTGAGCCACAGCTCGGGCGGTAGCACTGACCCCGCTCA 480

|||||

117 shIsAspHisAspGlu 123

|||||

481 TCATGACCACCATGATGAG 499

seq\_name: gb\_est1:AW536974

seq\_documentation\_block:

LOCUS AW536974 577 bp mRNA

DEFINITION G0110H10-3 NIA Mouse E7.5 Embryonic Portion cDNA Library Mus

musculus cDNA clone G0110H10 3', mRNA sequence.

ACCESSION AW536974

VERSION AW536974.1 GI:7179391

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 577)

AUTHORS Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac

, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood, W.H.

III, Becker, K.G. and Ko, M.S.H.

Genome-wide expression profiling of mid-gestation placenta and

embryo using a 15,000 mouse developmental cDNA microarray

Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)

20381348

CONTACT: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

Plate: G0110 row: H column: 10

Seq primer: -21M13 Forward

High quality sequence stop: 577

POLYA-Yes.

## FEATURES

source

Location/Qualifiers

1..577

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="G0110H10"

/clone\_lib="NIA Mouse E7.5 Embryonic Portion cDNA Library"

/sex="unknown"

/dev\_stage="7 5dpc Embryo"

/lab\_host="DHI08"

/note="Vector: pSPORT1 (Gibco/BRL Life Technology);

Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from

6 Embryo. The double-stranded cDNA was synthesized by

Gibco's kit with an oligo(dT) primer [NotI primer-adaptor

from GibcoBRL]

[5'-PGACTAGTTCTAGATCCGAGCGCCCTTTTCTTTTCTTTT-3']

from 0.5lug of mRNA. The double-stranded cDNAs were

treated with T4 DNA polymerase and purified by

ethanol-precipitation. The cDNAs were ligated to

Lone-linker Lt-Sal3 (include SalI sequence). The cDNAs

were purified by phenol/chloroform and separated from

free linkers by Centricon 100. Then, cDNAs were amplified

by long-range high fidelity PCR using Takara's Ex Taq

polymerase. Then, the cDNAs were purified by

phenol/chloroform and by Centricon 100. The cDNAs were

digested with SalI and NotI enzymes. Then, the cDNAs were

size selected by Gibco's Size Fractionation Column. The

cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid

vector. The DHI08 E. coli host was transformed with the

ligation mixture by chemical method. The library was

constructed by Xiaohong Wang and Minoru S. H. Ko.

137 a 160 c 126 g 154 t

alignment\_scores:

Quality: 632.00

Ratio: 5.402

Percent Similarity: 99.153

Percent Identity: 99.153

Length: 118

Gaps: 0

alignment\_block:

US-09-327-750D-30 x AW536974/rev

Align seg 1/1 to reverse of: AW536974 from: 1 to: 577

```

11 LeuAsnMetGluAsnAspHisGlnLysLysGluGluLysGluGluLysPr 27
|||||
576 CTCACATGGAGATGACCATCAGAAAAGAGGAGAGAGAGAGAGAGAGCC 527
|||||
27 oGlnAspThrIleArgArgGluProAlaValAlaLeuIleSerGluAlaG 44
|||||
526 ACAAGATACCATCAGAAGGAGGAGCAGCTGTGGCCCTGACCTCCGAGGCTG 477
|||||
44 lylLysAsnCysAlaProArgGlyGlyArgArgPheArgValArgGln 60
|||||
476 GCAAAACTGTGCACCTAGAGAGGTGCGAGGCGTTCCGGTTCGGCAG 427
|||||
61 ProIleAlaHisTyrArgTArgPLeuMetGlnArgValGlyGluProG1 77
|||||
426 CCCATGCTCATTATAGATGGGACCTGTATGCAGAGGGTTGGGAGGCCCA 377
|||||
77 nGlyArgMetArgGluGluAsnValGlnArgPheGlyGlyAspValArgG 94
|||||
376 GGAAGGATCAGAGAGAGAGAACCTACAGAGGTTTGGGGTGATGTGAGAC 327
|||||
94 lInLeuMetGluLysLeuArgGluArgGlnLeuSerHisSerLeuArgAla 110
|||||
326 AGCTCATGGAGAAGCTCAGGGAAGGAGCAGCTGAGCCACAGCCTGCGGGC 277
|||||
111 ValSerThrAspProProHisHisAspHisHisAspGluPheCysLeuMe 127
|||||
276 GTTAGCAGTGCAGCCGCTCATATGACCAACCATGATGAGTTTGCCTCAT 227
|||||
127 tPro 128
|||||
226 GCCC 223

```

seq\_name: gb\_est2:BF607762

```

seq_documentation_block:
LOCUS BF607762 785 bp mRNA EST 01-APR-2001
DEFINITION M11000704 Mouse 9-day fetus cDNA library ICRFp522 Mus musculus
CDNA clone ICRFp522A1947 5', mRNA sequence.
ACCESSION BF607762
VERSION BF607762.1 GI:13504254
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Yahyawi, M., Hennig, S., Neidhardt, L., Radelof, U., Hermann, B.G.,
Lehrach, H. and O'Brien, J.
Detection of a high number of novel genes in a 9-day mouse embryo
CDNA library normalised by oligonucleotide fingerprinting
Unpublished (2001)
Contact: Hennig S
Laborat 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inestr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
EST's are made from clones being representatives of clone clusters.
Clone clusters were calculated from oligonucleotide fingerprints.
PCR Primers
FORWARD: 5'-CAGCTATTCCAGAGTAGTGA-3'
BACKWARD: 5'-TAATAGACTCATATAGG-3'
Seq primer: 5'-ATTAGGTGACACTATAG-3'
High quality sequence stop: 785.
Location/Qualifiers
1..785
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="ICRFp522A1947"
/clone_lib="Mouse 9-day fetus cDNA library ICRFp522"

```

seq\_name: gb\_est2:BG228077

```

seq_documentation_block:
LOCUS BG228077 587 bp mRNA EST 08-FEB-2001
DEFINITION ux47c11.x1 Soares_NMAX_maxillary_process Mus musculus cDNA clone
IMAGE:3513237 3' similar to TR:09RLJ2 09RLJ2 BEX1 PROTEIN. ;, mRNA
sequence.
ACCESSION BG228077
VERSION BG228077.1 GI:12715592
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 587)
REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

```

```

/tissue_type="whole embryo"
/dev_host="embryonic 9-day"
/lab_host="E.coli, Xcl blue"
/note="Vector: PSVSPORT1; Site_1: NotI; Site_2: SalI;
Library preparation by oligo_dT priming of RNA. Clones can
be ordered from the Resource Center in Berlin,
http://www.rzpd.de."
BASE COUNT 190 a 188 c 233 g 153 t 21 others
ORIGIN

```

```

alignment_scores:
Quality: 591.00 Length: 128
Ratio: 4.844 Gaps: 4
Percent Similarity: 95.312 Percent Identity: 93.750

```

alignment\_block:

US-09-327-750D-30 x BF607762 ..

Align seg 1/1 to: BF607762 from: 1 to: 785

```

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHi 17
|||||
259 ATGGAGTCCAAAGATCAAGCGCTGAAATAATCTCAACATGGAGATGACCA 308
|||||
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArgG 34
|||||
309 TCAGAAAAGAGGAGAGAGAGAGAAAGCCACCAAGATACCATCAGAGGG 358
|||||
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProAr 50
|||||
359 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACCTGTGCACCTAG 408
|||||
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
409 GGAGGTGCGAGGCGTTCGGGTTCGGCAGCCCATCCCTCACTATAGATG 458
|||||
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGlu 84
|||||
459 GGACCTGATGCAGAGGTTGGGGAGGCCCCAGGAGAGGATGAGAGAGAGA 508
|||||
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuAr 100
|||||
509 ACGTACAGAGGTTTGGGGTGATGTGAGACAGATCATGGGGGAAGCTGAG 558
|||||
100 gGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp.ProPro 116
|||||
559 GGAAGGCCACCTGAACACAGAGCTGCGGGGTAGCACTGACCCGCCCT 608
|||||
117 His.HisAspHisHis, AspGluPhe 124
|||||
609 CATTATGACCAACCATGATGAGTTT 634

```

Email: cgapbs-r@mail.nih.gov  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:1393813  
High quality sequence stop: 439.

## FEATURES

source  
location/Qualifiers  
1..587  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3513237"  
/clone\_lib="Soares\_NMAX\_maxillary\_process"  
/tissue\_type="maxillary process"  
/lab\_host="DH10B (phage-resistant)"  
/note="vector: pT7T3-Pac (Pharmacia) with a modified  
polylinker; Site.1: NotI; Site.2: EcoRI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTTACCAATCTGAAGTGGGAGCGCGCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 138 a 160 c 121 g 168 t

## ORIGIN

alignment\_scores:

Quality: 586.00 Length: 117  
Ratio: 5.186 Gaps: 1  
Percent Similarity: 96.581 Percent Identity: 95.726

## alignment\_block:

US-09-327-750D-30 x BG228077/rev ..

Align seg 1/1 to reverse of: BG228077 from: 1 to: 587

13 MetGluAsnAspHisGlnLysGluGluLysGluGluLysProGlnAs 29  
587 ATGAGAATGACCATCAGAAAAAGGAGGAGAGGAAGTAATGCCACAAGA 538

29 pThrIleArgArgGluPro. AlavalAlaLeuIleSerGluAlaGlyLys 45  
537 TACCATFCAAGAGGAGCCAGGCTGTGGCCCTGACCTCCGAGGCTGGCAAA 488

46 AspCysAlaProArgGlyGlyArgArgPheArgValArgGlnProI 62  
487 AACTGTGCACCTTAGAGGAGGTCCGAGCGGTTCCGGGTTCCGAGCCCAT 438

62 eAlaHisTyrArgTrpAspLeuMetGlnArgValGlyGluProGlnGly 79  
437 CGCTCACTATAGATGGACCTGATGCAGAGGGTTGGGAGCCCCAGGGAA 388

79 rgMetArgGluGluAsnValGlnArgPheGlyGlyAspValArgGlnLeu 95  
387 GGATGAGAGAGGAGACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTC 338

96 MetGluLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSe 112  
337 ATGGAGAAGCTGAGGAAAGGAGCTGAGCCACAGCCTCGGGCGGTTAG 288

112 rThrAspProProHisHisAspHisHisAspGluPheCysLeuMetPro 128  
287 CACTGACCCGCTCATCATGACCAACCATGATGAGTTTTCCTCATGCC 239

seq\_name: gb\_est1:BE291071

## seq\_documentation\_block:

LOCUS BE291071 583 bp mRNA EST 13-JUL-2000  
DEFINITION 601086311F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3500522 5',  
mRNA sequence.  
ACCESSION BE291071  
VERSION BE291071.1 GI:9172545  
KEYWORDS EST.

## SOURCE

## ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 583)  
NIH-MGC http://mgi.nci.nih.gov/.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## CONTACT

## Tissue Procurement

## CDNA Library Preparation

## DNA Sequencing

## Clone distribution

## found through the

## http://image.llnl.gov

## Plate: LLAM8559

## High quality sequence start: 5

## High quality sequence stop: 503.

## FEATURES

## Location/Qualifiers

## Source

## 1..583

## /organism="Mus musculus"

## /strain="FVB/N"

## /db\_xref="taxon:10090"

## /clone="IMAGE:3500522"

## /clone\_lib="NCI\_CGAP\_Mam6"

## /sex="female, virgin"

## /tissue\_type="infiltrating ductal carcinoma"

## /dev\_stage="5 months"

## /lab\_host="DH10B"

## /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;

## Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.

## Library constructed by Life Technologies. Investigator

## providing samples: Jeffrey Green, M.D., NIH

## BASE COUNT 165 a 136 c 201 g 81 t

## ORIGIN

## alignment\_scores:

## Quality: 542.50 Length: 120

## Ratio: 4.801 Gaps: 2

## Percent Similarity: 94.167 Percent Identity: 90.000

## alignment\_block:

## US-09-327-750D-30 x BE291071 ..

## Align seg 1/1 to: BE291071 from: 1 to: 583

## 1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHi 17

## 208 ATGCAGTCCAAAGATCAAGCGCTGAAATAATCTCAACATGGAGATGACCA 257

## 17 sGlnLysLysGluGluLysGluGluLysProGlnAspThrIleArgArg 34

## 258 TCAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 307

## 34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50

## 308 AGCCAGCTGTGGCCCTGACCTCGAGGCTGGCAAAACTGTGCACCTAGA 357

## 51 GlyGlyValArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67

## 358 GGAGTGCAGAGCGGTTCCGGGTTCCGAGCCCATCTCCTCATATATG 407

## 67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluAla 84

## 408 GGACCTGATGCAGAGGTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 457

## 84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100

## 458 ACGTACAGAGGTTAGGGGGTGATGTGAGACAGCTCATGGAGAGCTGAG 507

```

101 GluArgGlnLeuSerHis...SerLeuArgAlaVal_SerThrAspProp 116
||||| ||||||| : : : ||||||| : : : ||||||| : : : |||||||
508 GAAGGCTGTGAGCCACAGACAGATCGAGCGGTTAGACAAATGACCCGA 557
||||| |||||||
116 roHisHis 118
|||||
558 CTCATCAT 565

seq_name: gb_est2:BF608209

* seq_documentation_block:
LOCUS BF608209 650 bp mRNA EST 01-APR-2001
DEFINITION MYL_001088 Mouse 9-day fetus cDNA library ICRFp522 Mus musculus
cDNA clone ICRFp522C0878 5', mRNA sequence.
ACCESSION BF608209
VERSION BF608209.1 GI:13504638
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 650)
Yahyawli, M., Hennig, S., Neidhardt, L., Radelof, U., Hermann, B.G.,
Lehrach, H. and O'Brien, J.
Detection of a high number of novel genes in a 9-day mouse embryo
cDNA library normalised by oligonucleotide fingerprinting
JOURNAL Unpublished (2001)
COMMENT Contact: Hennig S
Laboratory 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
EST's are made from clones being representatives of clone clusters.
Clone clusters were calculated from oligonucleotide fingerprints.
PCR Primers
FORWARD: 5'-GAGCTATTCCAGAACTAGTA-3'
BACKWARD: 5'-TAATACGACTCATTATAGG-3'
Seq primer: 5'-ATTAGTGACACTATAG-3'
High quality sequence stop: 650.
Location/Qualifiers
1..650
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="ICRFp522C0878"
/clone_lib="Mouse 9-day fetus cDNA library ICRFp522"
/tissue_type="whole embryo"
/dev_stage="embryonic 9-day"
/lab_host="E.coli, XL1 blue"
/note="Vector: PSVSPORT1; Site_1: NotI; Site_2: SalI;
Library preparation by oligo_dT priming of RNA. Clones can
be ordered from the Resource Center in Berlin,
http://www.rzpd.de."
BASE COUNT 164 a 167 c 180 g 128 t 11 others
ORIGIN

alignment_scores:
Quality: 537.00 Length: 128
Ratio: 4.629 Gaps: 0
Percent similarity: 90.625 Percent identity: 79.688

alignment_block:
US-09-327-750d-30 x BF608209
Align seq 1/1 to: BF608209 from: 1 to: 650

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
206 ATGGACTCCAAAGATCAAGCGGCGGAAATAATCTCAACATGGACACACCA 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 sGlnLysLysGluGluLysProGlnAspThrIleArgArg 34

```

```

|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
256 TCACAAAAGGACGACACAGCAAAAGCCACAGATACCATCTTAATGG 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
306 AGCAGCTGTGGCCCTTACCTCTCTAGCTAGCGCAAAAACCTGTGCACCTGA 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
356 GNAGGTGCGAGGCGGTTCCGGGTCGGCAGCCCATCCCTTACTATANATC 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluAla 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
406 GGACCTGATCCACAGGGTGGGGAGCCCCCATGGGAAGGATCAGACGACCA 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
456 ACGTACAGAGGCTCTGGGGGTGATGTGACAGAGCTCATGGAGAAACTCAGG 505
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPropH1 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
506 GAAAGGCAGCTTAGCCACAGCCTCGCGGCGGTAGCACCTGACNCGCCTCA 555
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
556 TCATGACTACCATGATTAGTTTTCCTCATGCCCC 589
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

seq_name: gb_gss:AZ936393

seq_documentation_block:
LOCUS AZ936393 637 bp DNA GSS 26-APR-2001
DEFINITION 2M0193L05F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0193L05 F, plasmid sequence.
ACCESSION AZ936393
VERSION AZ936393.1 GI:13794974
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 637)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenan, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.,
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0193 row: L column: 05
Seq primer: CGTTGTAAGACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 637.
Location/Qualifiers
1..637
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0193L05"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource

```







Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,  
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5'-GAGAGAGAGCGCGCAACTCGAGTTTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. cDNA went  
through one round of normalization to Rot = 5.0. Second strand cDNA  
was prepared with the primer adapter of sequence[5',  
GAGAGAGAGAGATCCAGAGTCAATTAAATTAATTAACCCCCCCCCC 3']. cDNA was  
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
XhoI. Host: SOLR.

FEATURES  
Location/Qualifiers  
1..480

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGD:MGI:1328321"  
/db\_xref="MGD:MGI:1900507"  
/clone="1110004J10"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="18 days embryo"  
<1..261  
/note="putative"  
/codon\_start=1  
/protein\_id="BAB22786.1"  
/db\_xref="GI:12834097"  
/translation="SGRNCAPRGRRRVRPROPIAHYRWDLMQRVGEPQGRMRRENVQ  
RFGDVRQLMEKRLERQLSHSLRAVSTDPHPHHHDEFCLMP"  
460..465  
/note="putative"  
480  
/note="putative"

CDS

polyA\_signal 117 a 103 c 132 g 128 t  
polyA\_site  
BASE COUNT 117 a 103 c 132 g 128 t  
ORIGIN

alignment\_scores:  
Quality: 471.00 Length: 86  
Ratio: 5.477 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.837

alignment\_block:

US-09-327-750D-30 x AK003429 ..

Align seg 1/1 to: AK003429 from: 1 to: 480

43 AlaGlyLysAsnCysAlaProArgGlyArgArgArgPheArgValAr 59  
:::|||||  
1 TCTGGCAAAACCTGTGCACCTAGAGGAGGTGCGAGCGGTTCCGGGTTG 50  
59 gGlnProIleAlaHisTyrArgTrpAspLeuMetGlnArgValGlyGluP 76  
|||||  
51 GCAGCCCATCGCTCACTATAGATGGACCTGATGCAGAGGGTTGGGGAGC 100  
76 roGlnGlyArgMetArgGluGluAsnValClnArgPheGlyGlyAspVal 92  
|||||  
101 CCCAGGGAGGATGAGAGAGGAGAACGTACAGAGGTTTGGGGGTGATGTG 150  
93 ArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSerHisSerLeuAr 109

|||||  
151 AGACAGCTCATGGAGAAAGCTGAGGGAAGGACGCTGAGCCACAGCCTGCG 200  
109 gAlaValSerThrAspProHisHisAspHisHisAspGluPheCysL 126  
|||||  
201 GCGGGTTAGCACTGACCCCGCCTCATCATGACCACCATGATGATTTTGCC 250  
126 euMetPro 128  
|||||  
251 TCATGCCC 258

**THIS PAGE BLANK (USPTO)**

OM of: US-09-327-750D-30 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Mar 11, 2002 3:50 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101154\_20340/app\_query.fasta\_1.1472  
-DB=Issued\_Patents\_NA -QPMF=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40 cdi -LISP=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=20000000000  
-USER=US09327750.@CGN1\_1\_99 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-327-750D-30

Query length: 128

Database: Issued\_Patents\_NA.\*

Database sequences: 351203

Database length: 113238999

Search time (sec): 146.090000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-381-881-1 +		93.50	169.00	1315	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-281-221-1 +		93.50	169.00	0.0919	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-931-999-4 +		86.50	137.69	5.09	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-809-740A-1 -		85.00	142.20	2.86	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-809-740A-4 -		85.00	142.20	2.86	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-426-599B-3 +		83.50	156.25	0.4714	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-506-553C-6 +		82.50	154.49	0.5910	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-995-1 +		82.50	146.08	1.74	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-403-852D-1 +		81.50	129.79	14.04	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-510-646B-1 +		81.50	129.79	14.04	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-231-818-1 +		81.50	129.79	14.04	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-705-1 +		81.00	141.68	3.05	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-393-985-17 +		81.00	138.85	4.39	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-951-115A-6 +		81.00	132.89	9.44	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-459-448A-6 +		81.00	132.89	9.44	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-595A-6 +		81.00	132.89	9.44	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-504B-6 +		81.00	132.89	9.44	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-444-6 +		81.00	132.89	9.44	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-053-549-7 +		81.00	132.89	9.44	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-717-294-42 +		79.50	127.68	18.40	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-053-549-1 +		79.00	119.52	52.37	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-147-777-3 +		78.50	107.74	237.26	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-452-872-3 +		78.50	107.74	237.26	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-660-645A-4 -		77.50	140.07	3.76	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-298-718-4 -		77.50	140.07	3.76	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-546-969-4 -		77.50	140.07	3.76	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-980-832-1 +		77.50	116.72	75.00	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-980-832-27 +		77.00	113.98	106.58	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-764-233A-1 +		77.00	97.60	870.99	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-452-075-1 -		76.50	126.19	22.27	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-231-061-1 -		76.50	126.19	22.27	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-402-282-3 -		76.50	108.49	215.70	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-508-004-3 +		76.50	108.49	215.70	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-402-066-3 +		76.50	108.49	215.70	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-402-068-3 -		76.50	108.49	215.70	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-592-874-1 -		76.50	102.17	484.98	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-096-942-2 -		76.50	102.17	484.98	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-096-867-2 -		76.50	102.17	484.98	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-258-261B-6 +		76.50	102.12	488.43	

/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-456-837-6 + 76.50 102.12 488.43  
/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-457-342-6 + 76.50 102.12 488.43  
/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-457-646A-6 + 76.50 102.12 488.43  
/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-458-076A-6 + 76.50 102.12 488.43  
/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-764-233A-4 + 76.50 102.12 488.43

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-381-881-1

seq\_documentation\_block:  
; Sequence 1, Application US/08381881  
; Patent No. 5939064  
; GENERAL INFORMATION:  
; APPLICANT: SAKELKOU, PAUL H., M.  
; APPLICANT: GAASTRA, WILLEM  
; TITLE OF INVENTION: BORDETTELLA BRONCHISEPTICA VACCINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STEVENS, DAVIS, MILLER & MOSHER  
; STREET: 515 NORTH WASHINGTON STREET.  
; CITY: ALEXANDRIA  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22314  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: POULOS III, JAMES A.  
; REGISTRATION NUMBER: 31,714  
; REFERENCE/DOCKET NUMBER: TPP 29685  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703 549-7200  
; TELEFAX: 703 528-5313  
; TELEX: 44-0704 "STEVENS"  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1315 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bordetella bronchiseptica  
; STRAIN: 401  
; IMMEDIATE SOURCE:  
; CLONE: E coli PC2495(pIVB3-420)  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1..539  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 540..1142  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1143..1315  
; US-08-381-881-1

alignment\_scores:  
Quality: 93.50 Length: 96  
Ratio: 1.908 Gaps: 4  
Percent Similarity: 51.042 Percent Identity: 33.333  
alignment\_block:  
US-09-327-750D-30 x US-08-381-881-1 ..

```

Align seg 1/1 to: US-08-381-881-1 from: 1 to: 1315
43 AlaGlyLysAsnCysAlaProArgGlyGlyArgArgPheArgValar 59
||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
542 GCAAGCCAAACAGTCTCTCTGGCGCGCGCTCGCGCGCTCG 591
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
59 gGlnProIle.....AlaHisTyrArgTrp..... 67
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
592 CCGCCCATCGGAGAGCGCACCATTCATACCGCGCAGCACCGAC 641
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
68 ..AspLeuMetGlnArgValGlyGluPro.....GlnGlyArgMet 80
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
642 CAGACCTGCACGATCGAGGAGCCCGCGGTTACATCAAGGTGCGCA 691
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
81 ArgGluGluAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetG1 97
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
692 CCTGCCAC.....GATCTCCAAGAGCGCGCTGA 720
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
97 uLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThra 114
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
721 AGAAGCGCGGCGAGCTGGCGGGCGCACTCGCTTCGATATCAAGCTGAAG 770
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
114 spProHisHisAspHisAspGluPheCysLeu 126
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
771 GACTGCCCGACCGTCAACACTCTCAAGCTGTACTT 808
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-281-221-1

seq_documentation_block:
; Sequence 1, Application US/09281221
; Patent No. 6284256
; GENERAL INFORMATION:
; APPLICANT: SAKELAKOUL, PAUL H.,M.
; TITLE OF INVENTION: BORDETTELLA BRONCHISEPTICA VACCINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STEVENS, DAVIS, MILLER & MOSHER
; STREET: 515 NORTH WASHINGTON STREET.
; CITY: ALEXANDRIA
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/281,221
; FILING DATE: 30-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/381,881
; ATTORNEY/AGENT INFORMATION:
; NAME: POULOS III, JAMES A.
; REGISTRATION NUMBER: 31,714
; REFERENCE/DOCKET NUMBER: TPP 29685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 549-7200
; TELEFAX: 703 528-5313
; TELEX: 44-0704 "STEVENS"
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

```

; ORIGINAL SOURCE:
; ORGANISM: Bordetella bronchiseptica
; STRAIN: 401
; IMMEDIATE SOURCE:
; CLONE: E coli PC2495(pIVB3-420)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..539
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 540..1142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1143..1315
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-281-221-1

alignment_scores:
Quality: 93.50 Length: 96
Ratio: 1.908 Gaps: 4
Percent Similarity: 51.042 Percent Identity: 33.333

alignment_block:
US-09-327-750D-30 x US-09-281-221-1 ..
Align seg 1/1 to: US-09-281-221-1 from: 1 to: 1315
43 AlaGlyLysAsnCysAlaProArgGlyGlyArgArgPheArgValar 59
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
542 GCAAGCCAAACAGTCTCTCTGGCGCGCGCTCGCGCGCTCG 591
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
59 gGlnProIle.....AlaHisTyrArgTrp..... 67
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
592 CCGCCCATCGGAGAGCGCACCATTCATACCGCGCAGCACCGAC 641
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
68 ..AspLeuMetGlnArgValGlyGluPro.....GlnGlyArgMet 80
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
642 CAGACCTGCACGATCGAGGAGCCCGCGGTTACATCAAGGTGCGCA 691
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
81 ArgGluGluAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetG1 97
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
692 CCTGCCAC.....GATCTCCAAGAGCGCGCTGA 720
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
97 uLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThra 114
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
721 AGAAGCGCGGCGAGCTGGCGGGCGCACTCGCTTCGATATCAAGCTGAAG 770
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
114 spProHisHisAspHisAspGluPheCysLeu 126
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||
771 GACTGCCCGACCGTCAACACTCTCAAGCTGTACTT 808
||||:||||| ||| ||| ||| ||| ||| ||| ||| |||

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-931-999-4

seq_documentation_block:
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Tandolo, John J.
; APPLICANT: Crupper, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```



```

; OTHER INFORMATION: /note= "ORF2"
; US-08-809-740A-4

alignment_scores:
    Quality: 85.00      Length: 127
    Ratio: 1.197      Gaps: 6
    Percent Similarity: 55.906      Percent Identity: 29.921

alignment_block:
US-09-327-750D-30 x US-08-809-740A-4/rev ..

Align seg 1/1 to reverse of: US-08-809-740A-4 from: 1 to: 3252

16 AspHisGlnLysLysGluGluLysProGlnAspThr...IL 31
:::||||| :::: :::: :::: :::: :::: :::: :::: ::::
2772 GAGCACCAAGTCAGATCGTACTCCGGCGCGCGGGAGACGTTGACCG 2723
31 eArgArgGluProAlaValAlaLeuIleSerGlu..... 42
||||| :::: :::: :::: :::: :::: :::: :::: ::::
2722 TCGCGGGAGCAGCGCCCTCGCGAGCGCCAGCAGCGCGTCGCCACGTCG 2673
43 .....AlaGlyLysAsnCysAlaProArgGlyArgArgArg 55
||| :::: :::: :::: :::: :::: :::: :::: ::::
2672 AGGGCGCGCGCGCGCGTGTAGCGGGCGGTCTATGCTCTTCGGCGCGGT 2623
56 PheArg.....ValArgGlnProIleAlaHisTyrArgTrpAspLeuMe 70
||| :::: :::: :::: :::: :::: :::: :::: ::::
2622 CACCGGAACGCGTACGGTCCGAGCAGCGCGCGCGTCGGCGCTCCG 2573
70 tGlnArgValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnA 87
::: :::: :::: :::: :::: :::: :::: :::: ::::
2572 CCGCGTCTCGTCCGGGTCGCCGAGCGGTCCGGAACACCGCTCGACC 2523
87 rpPheGlyGlyAspValArgGlnLeuMetGlnLysLeuArg...GluArg 102
||| :::: :::: :::: :::: :::: :::: :::: ::::
2522 GCCTCGGTCCGACCTCCGCTCGTCCAGTCCGCTCGCGCGCGCGCAG 2473
103 GlnLeuSerHisSerLeuArg.....AlaVal 111
||| :::: :::: :::: :::: :::: :::: :::: ::::
2472 CAGTTTCGGCGCGCGCGCCGACCGCGCGGGGGTCCGAATGTGGCGCGGT 2423
111 lSerThrAsp.....ProProHisHis 118
||| :::: :::: :::: :::: :::: :::: :::: ::::
2422 ACCCATCGATCGCGCGTACCACCGCGCTCAC 2392

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-426-599B-3

seq_documentation_block:
; Sequence 3, Application US/08426599B
; Patent No. 5932438
; GENERAL INFORMATION:
; APPLICANT: Uriach-Marsal, Juan
; APPLICANT: Rubio-Susan, Victor
; APPLICANT: Patino-Martin, Cristina
; APPLICANT: Kalo-Koenova, Eliza I.
; APPLICANT: del Moral-Juarez, Catalina
; APPLICANT: Faus-Santassana, Ignacio
; APPLICANT: del Rio-Pericacho, Jose-Luis
; APPLICANT: Blade-Pique, Joan
; TITLE OF INVENTION: Preparation of Thaumatin Sweeteners
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701 East
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,599B
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Stephen A.
; REGISTRATION NUMBER: 38,609
; REFERENCE/DOCKET NUMBER: 1604-123A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-624-1589
; TELEFAX: 202-783-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Optimized cDNA"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..621
; US-08-426-599B-3

```

```

alignment_scores:
  Quality: 83.50      Length: 96
  Ratio: 1.670       Gaps: 6
  Percent Similarity: 52.083      Percent Identity: 34.375

```

```

alignment_block:
US-09-327-750D-30 x US-08-426-599B-3

```

Align seg 1/1 to: US-08-426-599B-3 from: 1 to: 624

```

33 ArgGluProAlaValAlaLeuSerGluAlaGlyLysAsnCysAlaPr 49
||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
15 CGTCAACG.....CTGCTCTACACCGTCTGGCGCGCGCTCCNA.. 56

49 oArgGlyArgArgPheArgValArgGlnProIleAlaHisTyrA 66
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57 .....GGCGACGGCGCTCGACCGCGCGCGCGCGCGCTCAACTCC 99

66 rGrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGlu 82
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
100 GG.....CGAGTCTGGACCATCAACGTCGAGCCCGGCAC 134

83 GluAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLe 99
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 CAGAGCGCGGCAAGTCTGGCGCGCGCAC...CGACTGCTACTTCGAGCACT 181

99 uArgGluArgGlnLeuSerHis..... 106
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 CCGCGCGCGCATCTGCGGACCGCGGCGGCGGCGCTCTCCAGTGC 231

107 ..SerLeuArgAlaValSerThrAspProHisHis 118
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 AAGCGCTTCG.....CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 257

```

seq\_name: /cqn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-506-553C-6

```

seq_documentation_block:
; Sequence 6, Application US/08506553C
; Patent No. 6120989
; GENERAL INFORMATION:
; APPLICANT: Vornhagen, Rolf; Hinderer, Walter; Someborn, Han-H.;
; APPLICANT: Plancher, Bodo; and Jahn, Gerhard
; TITLE OF INVENTION: ISOLATED HUMAN CYTOMEGALOVIRUS
; TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF

```

```

; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,553C
; FILING DATE: 07/25/95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LEDER 202-PFF/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-506-553C-6

```

```

alignment_scores:
  Quality: 82.50      Length: 117
  Ratio: 1.650       Gaps: 4
  Percent Similarity: 42.735      Percent Identity: 26.496

```

alignment\_block:

US-09-327-750D-30 x US-08-506-553C-6

Align seg 1/1 to: US-08-506-553C-6 from: 1 to: 607

```

50 ArgGlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrA 66
||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 CGGTCTGGAGAGGTGGCGCGGTGGCGGCAAGAA...GCACGACGCG 153

66 gTrpAspLeuMetGlnArgValGlyGluProGln..... 77
|||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 GTGGCGCGGTGTTCGGTACGCGGAAATAGTAGCTGCGTGGCGCGGC 203

77 ..... 77

204 GGTGATCATCACCGGTCTTCTCTCAAGAAATACGACGACGACAA 253

78 .....GlyArgMetArgGluG1 83
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 GATCACCAGCTACCTGACGTCCAAAGGTGATCGGCGCGCGGAGGAG 303

83 uAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuA 100
||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 GAGGAGCGCGCGTTGGATCGCAACTCCGCAATTACTTCAACGACGCG 353

100 rg.....GluArgGlnLeuSer..... 105
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 AAGAGGAGAGCGACACCGAGGATCTGTACGTTCCAGTTCGTCCTAA 403

106 HisSerLeuArgAlaValSerThrAspProHisHisAspHisHis 122
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

404 CACCAAGAAGCAAAAGTCGGCGAAGATCCTGAAGAGCACGACGCGCATGA 453  
122 p 122  
454 A 454

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-030-995-1

seq\_documentation\_block:

; Sequence 1, Application US/09030995  
; Patent No. 6235519  
; GENERAL INFORMATION:  
; APPLICANT: Blanc, Veronique  
; APPLICANT: Blanc, Francis  
; APPLICANT: Crouzet, Joel  
; APPLICANT: Jacques, Nathalie  
; APPLICANT: Lacroix, Patricia  
; APPLICANT: Thibaut, Denis  
; APPLICANT: Zagorec, Monique  
; APPLICANT: Debussche, Laurent  
; APPLICANT: De Crecy-Lagard, Valerie  
; TITLE OF INVENTION: Polypeptides Involved In The  
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flanagan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/403,852D  
; FILING DATE: 10-MAY-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR 93/00923  
; FILING DATE: 25-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 92/11441  
; FILING DATE: 25-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 03806.0054-00000  
; TELEPHONE: (202) 408-4000  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5392 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: S.pristinaespiralis  
; US-08-403-852D-1

alignment\_scores:  
Quality: 82.50 Length: 123  
Ratio: 1.231 Gaps: 8  
Percent Similarity: 54.472 Percent Identity: 30.894

alignment\_block:

US-09-327-750D-30 x US-09-030-995-1 ..  
Align seg 1/1 to: US-09-030-995-1 from: 1 to: 1365  
27 ProGlnAspThrIleArgArg...GluProAlaValAlaLeuIleSerG1 42  
908 CTTATGTCACAAATCCGCGCGTGTCTACATCGGCATCGACGACGATGG 957  
42 uAlaGlyLysAsnCysAla...ProArgGlyGlyArgArg..... 55  
958 TTCACCTTCACATGTCGACGCCCGCGCTGTGGCGCGCGACGTGAT 1007  
56 .....PheArgValArgGlnProIleAlaHisTyArgTirAspLeuMet 70  
1008 CCTGGCGCGCATCGCCCTGCCCGCGGCAAGCAGGAGATGTCGCCGATG 1057  
71 GlnArgValGlyGlu.....Pro.GlnGlyArgMetArgGluG 83  
1058 TCGAGGCCCGCGTCCGCCCGCGGAGACCGCGCAAGGACGACTACGACGCG 1107  
83 luAsnValGln.....ArgPheGlyGlyAspValArgGlnLeuMet 96  
1108 ATCCGCTATCAGGGCGGACTAGTCAAGAGCTGATCCCGGAGACCGACTA 1157  
97 GluLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerTh 113  
1158 TCCGAGCTTCGACGTCGACGCGGCCAACGAAGCCTTCTTCGAGTGAAGA 1207  
113 rAsp.....ProPro..... 116  
1208 AGCACAGAAGAAAGACATCATGAGTTCCGCCACACACTCCTATCGGTCG 1257  
117 ..HisHisAspHisHis 121  
1258 GTCATCACCGGCACCAT 1274  
seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-403-852D-1

seq\_documentation\_block:  
; Sequence 1, Application US/08403852D  
; Patent No. 5891655  
; GENERAL INFORMATION:  
; APPLICANT: Blanc, Veronique  
; APPLICANT: Blanc, Francis  
; APPLICANT: Crouzet, Joel  
; APPLICANT: Jacques, Nathalie  
; APPLICANT: Lacroix, Patricia  
; APPLICANT: Thibaut, Denis  
; APPLICANT: Zagorec, Monique  
; APPLICANT: Debussche, Laurent  
; APPLICANT: De Crecy-Lagard, Valerie  
; TITLE OF INVENTION: Polypeptides Involved In The  
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flanagan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/403,852D  
; FILING DATE: 10-MAY-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR 93/00923  
; FILING DATE: 25-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 92/11441  
; FILING DATE: 25-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 03806.0054-00000  
; TELEPHONE: (202) 408-4000  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5392 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: S.pristinaespiralis  
; US-08-403-852D-1

alignment\_scores:  
Quality: 81.50 Length: 118  
Ratio: 1.430 Gaps: 5  
Percent Similarity: 48.305 Percent Identity: 27.966

alignment\_block:

US-09-327-750D-30 x US-08-403-852D-1 ..  
Align seg 1/1 to: US-08-403-852D-1 from: 1 to: 5392  
17 HisGlnLysLysGluGluLysGluLysProGlnAspThrIleArgAr 33  
4800 CATCGTGTCACCGGTTCATCGCCACCGACCATCTGATGGTCTTCCCGC 4849







```

981 AGAAAAAGGTATACATATGATAAATGGGAAAGTTACACAGAAGAAATG 1030
      ::::: ::::: ::::: ::::: ::::: :::::
78  lYArgMetArgGluGluAsnValGlnArgPheGlyGlyAspValArgGln 94
      ::::: ::::: ::::: ::::: ::::: :::::
1031 AAGAAATGGAGGAAACAGTGTGCCAGCAT.....GGGAGAGTACATGAG 1074
      ::::: ::::: ::::: ::::: ::::: :::::
95  LeuMetGluLysLeuArgGluArgGlnLeuSerHisSerLeuArgAlaVa 111
      ::::: ::::: ::::: ::::: ::::: :::::
1075 ACGATG...AAGCAAGGCTAAGGCAGCTGGATAAGCACAGCAGGCCAC 1121
      ::::: :::::
111  lSer 112
      :::::
1122 AGCC 1125

```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-393-985-17

```

seq_documentation_block:
; Sequence 17, Application US/08393985
; Patent No. 5693476
; GENERAL INFORMATION:
; APPLICANT: Scheller, Richard H.
; TITLE OF INVENTION: Methods and Compositions for Modulation
;   OF INVENTION: of Vesicular Release
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,985
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 8600-0152
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2040 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Mouse SNAP-25 (GenBank M22012)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 164..784
US-08-393-985-17

```

```

alignment_scores:
  Quality: 81.00      Length: 125
  Ratio: 1.110      Gaps: 6
Percent Similarity: 58.400      Percent Identity: 25.600

```

alignment\_block:  
US-09-327-750d-30 x US-08-393-985-17

```

Align seg 1/1 to: US-08-393-985-17 from: 1 to: 2040

2  GluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHisG1 18
      ||||| ||||| ||||| ::::: ::::: :::::
275 GAGAGTAAAGATCTGTCATCAGGACTTTGGTTATGTTGATGAGCAAGG 324
      ||||| ||||| ||||| ::::: ::::: :::::
18  nLysLysGluGluGluGluLysProGlnAspThrIleArgArgGluP 35
      ||||| ||||| ||||| ::::: ::::: :::::
325 CGAACAACCTGGAACGATTCAGGAGGAGGATGGACCAATCAATAAGAT. 373
      ||||| ||||| ||||| ::::: :::::
35  roAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArgGly 51
      ::::: ::::: ::::: ::::: :::::
374 .....ATGAAAGAACGACAAAGAAATTGACGACCTAGGA 409
      ::::: ::::: ::::: :::::
52  GlyArgArgArgPhe.....ArgValar 59
      ::::: :::::
410 .....AAATTCGCGGGCTTTGTGTGTCTGTCCTGTAAACAGCTTAA 450
      ::::: :::::
59  gGlnProIleAlaHisTyrArg...TrrAspLeuMetGlnArg.....V 73
      ::::: :::::
451 ATCCAGTGATGCTTACAAAAAGCCTGGGCAATAATCAGGATGGAGTAG 500
      ::::: :::::
73  aGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArg..... 87
      ||||| ::::: ::::: :::::
501 TGGCCAGCCAGCCTGCCCTGTGTGTGTGATGACGGGAGCAGATGCCATC 550
      ||||| ::::: :::::
88  PheGlyGlyAspValArgGlnLeuMetGluLysLeuArgGluGlnLe 104
      ||||| ::::: :::::
551 AGTGGTGGCTTCATCCGAGGGTAACAACGATGCCCGGGAATAATGAAT 600
      ::::: :::::
104 userHisserLeuArgAlaValSer 112
      ::::: :::::
601 GGATGAAACCTAGACAGCGTGAGC 625

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-951-715A-6

seq_documentation_block:
; Sequence 6, Application US/07951715A
; Patent No. 5625136
; GENERAL INFORMATION:
; APPLICANT: Koziele, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merilin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;   CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/951,715A

```

```

: FILING DATE: 25-SEP-1992
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/772,027
: FILING DATE: 04-OCT-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Spruill, W. Murray
: REGISTRATION NUMBER: 32,943
: REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
: TELEPHONE: (919)541-8615
: TELEFAX: (919)541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3624 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Synthetic DNA"
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3621
: OTHER INFORMATION: /product= "Full-length, maize
: OTHER INFORMATION: optimized cryiB"
: OTHER INFORMATION: /note= "Disclosed in Figure 6."
US-07-951-715A-6

alignment_scores:
  Quality: 81.00 Length: 121
  Ratio: 1.397 Gaps: 4
  Percent Similarity: 47.934 Percent Identity: 23.967

alignment_block:
US-09-327-750D-30 x US-07-951-715A-6 ..
Align seg 1/1 to: US-07-951-715A-6 from: 1 to: 3624

14 GluAsnAspHisGlnIysLysGluGluLysGluLysProGlnAspTh 30
||||:||||| :||||:||||| :||||:|||||
1026 GAACACGAGCAGCCAGCGCGCCACCAACAGCAGCATCAACCGGTGACCC 1075
30 rIleArgArgGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnC 47
||||:||||| :||||:||||| :||||:|||||
1076 TGGCTTCGCGCGCGGCGGTGTACCGCACCGAGAGCTACGCGGCGTG 1125
47 ysAla.....ProArgGlyGlyArgArgArg 55
:||||| :||||| :||||:|||||
1126 CTGCTGGGGGCATCTACCTGGAGCCCATCCAGCGGTGCCACCGTGG 1175
56 PheArgValArgGlnProIleAlaHisTyrArgTrpAspLeuMetGlnAr 72
:||||:||||| :||||:||||| :||||:|||||
1176 CTTCAACTTCACCAACCCCGCAGACATCATCGCGA.....CCGCG 1213
72 gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
| :||| :||| :||||:|||||
1214 GCACCGCCAACTACAGCAGCCCTACGAGAGCC..... 1247
89 lyGlyAspValArgGlnLeuMetGluLysLeuArgGluArg..... 102
1248 .....CGCGCTGCAGCTGAAGGACAGCAGCAGCTGCCGCC 1286
103 .....GlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
: : : : : : : : : : : : : : : : :
1287 CGAGACACGAGCGCGCCCAACTACGAGAGCTACAGCCACCGCTGAGCC 1336
117 sHisAspHisHis 121
||| |||||
1337 ACATCGGCATCAT 1349

```

seq\_name: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq:US-08-459-448A-6

```

seq_documentation_block:
: Sequence 6, Application US/08459448A
: Patent No. 5859336
: GENERAL INFORMATION:
: APPLICANT: Koziel, Michael G.
: APPLICANT: Desai, Nalini M.
: APPLICANT: Lewis, Kelly S.
: APPLICANT: Kramer, Vance C.
: APPLICANT: Warren, Gregory W.
: APPLICANT: Evoia, Stephen V.
: APPLICANT: Crossland, Lyle D.
: APPLICANT: Wright, Martha S.
: APPLICANT: Merlin, Ellis J.
: APPLICANT: Launis, Karen L.
: APPLICANT: Rothstein, Steven J.
: APPLICANT: Bowman, Cindy G.
: APPLICANT: Dawson, John L.
: APPLICANT: Dunder, Erik M.
: APPLICANT: Pace, Gary M.
: APPLICANT: Suttie, Janet L.
: TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
: NUMBER OF SEQUENCES: 94
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5859336artis Corporation
: STREET: Patent & Trademark Dept., 520 White Plains
: STREET: Rd., POB 2005
: CITY: Tarrytown
: STATE: New York
: COUNTRY: USA
: ZIP: 10591-9005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/459,448A
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/951,715
: FILING DATE: 25-SEP-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/772,027
: FILING DATE: 04-OCT-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Pace, Gary M.
: REGISTRATION NUMBER: 40403
: REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919)541-8582
: TELEFAX: (919)541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3624 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Synthetic DNA"
: HYPOTHETICAL: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3621
: OTHER INFORMATION: /product= "Full-length, maize
: OTHER INFORMATION: optimized cryiB"
: OTHER INFORMATION: /note= "Disclosed in Figure 6."
US-08-459-448A-6

```

alignment\_scores:

```

33  rgGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaPro 49
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
264  GGGAGCCC...TTGGCCCTCCCTTTGGATGCTGCTGAATACTGTGTGCT 310
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
50  ArgGlyGlyArgArgGpbeArgValGlnProGlnProIleAlaHisItyrAr 66
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
311  AGAGGAANTCGTAGGCGGTTCCGCGTTAGCAGCACCCTCTGCAGTATAG 360
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
66  gTTPAsPleuMetGlnArgValGlyGluProGlnGlyArgMetArgGluG 83
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
361  ATGGCATATGATCATAGGCTTGGAGAACACACAGGCAAGGATGAGAGAAG 410
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
83  luAsnValGlnArgGpbeGlyClyAspValArgGlnLeuMetGluLysLeu 99
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
411  AGAATATGGAAGAGATTGGGAGGGGTGACACAGCTGATGGAAGACGTG 460
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
100  ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
461  AGGCAAAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCACATGACCCCC 510
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
116  oHisHisAspHisHisAspGluPhecysLeuMetPro 128
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
511  TCACCATGACCATCATGATGATGTTTGCTTTATGCCC 547
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT :AAH75810
seq_documentation_block:
ID   AAH75810 standard; cDNA; 767 BP.
XX
XX
AC   AAH75810;
XX
XX
DT   17-OCT-2001 (first entry)

```

```

alignment_block:
  US-09-327-750D-30 x AAH75810  ..
  Align seg 1/1 to: AAH75810 from: 1 to: 767

1 MetGluSerLysAspGln...GlyValLysAsnLeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 ATGGAGTCCAAAGAGGAACGAGCGTTAAACAATCTCATCTGGGAAATGT 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16 pHisGlnLysLysGluGluLysGluGluLysProGlnAspThrIleArgA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 CAACCAAGGAAAATGATCAAAAAGATGAAAGGACCAAGTTCCTAATAAG 200
|||||:|||||:|||||:|||||:|||||:|||||:
33 rgGluProAlaValAlaLeuIleSerGluAlaGlaGlyLysAsnCyAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 GGGAGGCC...TTGGGCCCTACCTTTGAATGTAGTGAATCTGTGTGGCT 247
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 AtgGlyGlyVarGArgPheArgValArgGlnProIleAlaHisTyrAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 AGAGGAAACCGTAGCGGTTCCCGCTTAGCGAGCCCATCTCGCATATAG 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 gTrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluG 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 ATGGGACATAATGCATAGGCTTGGAGAGCCACAGCAAGGATGAGAGAGG 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 AGAATATGGAAGAGATTGGGAGGAGGTGAGACAGCTGATGGAAAGCTG 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
398 AGSGAAAAGCAGTTGAGTCATAGTCTCGGGSCAGTCAGCACTGATCCCCC 447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
448 TCACCATGACCATCAGCATGAGTTTGGCTTATGCC 484
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq name: /SID52/acadata/geneseq/geneseq/NA2001.DAT:AAI58581

```

alignment_scores:		
Quality:	450.00	Length: 129
Ratio:	3.879	Gaps: 2
Percent Similarity:	89.922	Percent Identity: 66.667

OM of: US-09-327-750D-30 to: N\_Geneseq\_1101.\* out\_format : pfs  
Date: Mar 11, 2002 3:42 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=framet.p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101154\_20362/app\_query.fasta\_1.1472  
-DB=N\_Geneseq\_1101 -Qfmt=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -GAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US09327750.@CGL\_1\_330 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-327-750D-30  
Query length: 128  
Database: N\_Geneseq\_1101.\*  
Database sequences: 930621  
Database length: 428662619  
Search time (sec): 355.560000

score\_list:

Sequence	Strd Orig	zScore	EScore Len	Documentation
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH45143 +		450.50	857.22	1
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH75810 +		450.00	856.56	792
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAI158581 +		450.00	855.45	767
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAF59611 +		450.00	855.06	862
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAI03880 +		437.00	832.87	898
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAI03879 +		437.00	832.36	662
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAI0367 +		437.00	830.40	698
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAI0367 +		437.00	830.40	858
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAI0367 +		437.00	830.40	891
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH03517 +		171.50	317.52	891
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH03517 +		170.00	314.90	865
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH13750 +		170.00	311.56	1229
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH85548 +		170.00	310.56	1364
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAF23528 +		160.00	297.62	700
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAF21748 +		143.50	263.19	917
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAI031005 +		119.50	222.05	532
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH87147 +		100.50	195.34	0.0084
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH06183 +		99.00	184.71	0.0327
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH06183 +		99.00	183.93	421
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAI03004 +		93.50	172.74	0.1317
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH09292 +		93.50	167.46	0.2985
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH06184 +		91.50	169.58	0.2277
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH16713 +		91.50	155.07	1.46
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH82586 +		88.50	164.33	0.4464
/SID22/gcgdata/geneseq/geneseq/NA1997.DAT:AAV74547 +		87.50	135.09	18.98
/SID22/gcgdata/geneseq/geneseq/NA1998.DAT:AAV21511 +		86.50	134.13	21.46
/SID22/gcgdata/geneseq/geneseq/NA1998.DAT:AAH14635 +		86.00	140.28	9.75
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH75924 +		86.00	136.57	15.69
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ00461 +		85.50	146.85	4.20
/SID22/gcgdata/geneseq/geneseq/NA1998.DAT:AAV61594 +		85.00	151.70	2.26
/SID22/gcgdata/geneseq/geneseq/NA1996.DAT:AAH25480 +		85.00	138.20	12.73
/SID22/gcgdata/geneseq/geneseq/NA1996.DAT:AAH06312 +		83.50	151.04	2.45
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH87834 +		83.00	136.91	5.62
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH11192 +		83.00	151.07	15.03
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH65399 +		82.50	149.48	3.00
/SID22/gcgdata/geneseq/geneseq/NA1995.DAT:AAI03508 +		82.50	149.37	3.04
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH21172 +		82.50	141.65	8.18
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH63400 +		82.50	136.82	15.20
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH68525 +		82.50	88.79	7.2e+03
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH47167 +		82.00	131.21	31.20

/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ53884 + 81.50 143.82 6.20  
/SID22/gcgdata/geneseq/geneseq/NA1994.DAT:AAQ64201 + 81.50 126.63 56.18  
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH40054 + 81.00 143.24 6.67  
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH00370 + 81.00 142.48 7.35  
/SID22/gcgdata/geneseq/geneseq/NA1998.DAT:AAV57358 + 81.00 134.92 19.38

seq\_name: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH45143

seq\_documentation\_block:

ID AAH45143 standard; cDNA; 792 BP.

XX AAH45143;

DT 07-SEP-2001 (first entry)

DE Human brain expressed X-linked protein, hBex, coding sequence.

XX Human; brain expressed X-linked protein; cytostatic; auditory; nontropic;  
KW hBex; dysembryoplasia; hereditary disease; cancer; tumour; deafness;  
KW X-chromosome-binding mental retardation; lissencephalous disease; ss.

XX Homo sapiens.

XX WO200140286-A1.

PD 07-JUN-2001.

PF 27-NOV-2000; 2000WO-CN00502.

XX 30-NOV-1999; 99CN-0124179.

PA (BIOR-) BIOROAD GENE DEV LTD SHANGHAI.

PI Mao Y, Xie V;

DR WPI; 2001-397944/42.

XX P-PSDB; AAB99224.

PT Isolated human brain-expressed X-linked polypeptide used to diagnose  
PT and treat of dysembryoplasia, hereditary diseases, cancer, tumor,  
PT deafness and X-chromosome-binding mental retardation -

PS Claim 5; Page 22; 30pp; Chinese.

CC The present sequence is the coding sequence for a human brain-expressed  
CC X-linked protein (hBex). hBex and its coding sequence are useful in the  
CC diagnosis and treatment of dysembryoplasia, hereditary diseases, cancer,  
CC tumours, deafness, X-chromosome-binding mental retardation and  
CC lissencephalous disease. hBex is also useful for screening mimics,  
CC agonists, or inhibitors, and in peptide fingerprinting identification.  
CC hBex coding sequence can be used as primers or probes, or in producing  
CC gene chips or microarrays.

SQ Sequence 792 BP; 214 A; 172 C; 219 G; 187 T; 0 other;

alignment\_scores:

Quality: 450.50 Length: 129  
Ratio: 3.952 Gaps: 3  
Percent Similarity: 88.372 Percent Identity: 68.992

alignment\_block:

US-09-327-750D-30 x AAH45143 ..

Align seg 1/1 to: AAH45143 from: 1 to: 792

1 MetGluSerLysAspGln...GlyValLysAsnLeuAsnMetGluAsnAs 16

173 ATGGAGTCCAAAGAAACTAGCAGTAAACAGTCTCAGCATGCAAAATGC 222

16 pHISGlnLysLysGluGluGluLysProGlnAspThrIleArga 33

223 CAACCAAGAAATGAAGAAAGGAG.....CAAGTTGCTTAATAAAG 263







```
; APPLICATION NUMBER: US/08/692,922
; FILING DATE: 31-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Shaw, Melissa A.
;   REGISTRATION NUMBER: 38,301
;   REFERENCE/DOCKET NUMBER: PM-8935
;   TELEPHONE: 414-273-2100
;   TELEFAX: 414-223-5000
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 5917 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: join(942..1085, 3400..3753)
;   US-08-692-922-1

Query Match
Best Local Similarity 5.0%; Score 34.8; DB 4; Length 5917;
Matches 60; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 176 catggccaatgcccacagaaacgaagagctggagcagccctcagaatggacagga 235
DB 3748 CAGGGCTGAGGTCACCTCGCCGAGCATGTCCTGCAGACCCCTCGAGCCTGCTCAGGG 3689

QY 236 acaccgcccctggaggagggtgagggccaccagcctgtgc 277
DB 3688 CCACCACCTCGGTAGTAGAGGAGGCTTCAGGAGCGCGC 3647

RESULT 12
US-09-178-115-110/c
; Sequence 110, Application US/09178115
; Patent No. 6297041
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; FILE REFERENCE: D-0021.5A
; CURRENT APPLICATION NUMBER: US/09/178,115
; EARLIER FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 09/177,776
; EARLIER FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 08/787,739
; EARLIER FILING DATE: 1997-01-24
; EARLIER APPLICATION NUMBER: 08/485,049
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/486,756
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/477,504
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/481,658
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485,862
; EARLIER FILING DATE: 1994-06-15
; EARLIER APPLICATION NUMBER: 08/177,093
; EARLIER FILING DATE: 1993-12-30
; EARLIER APPLICATION NUMBER: 07/964,589
; EARLIER FILING DATE: 1992-10-21
; EARLIER APPLICATION NUMBER: PV-709-92
; EARLIER FILING DATE: 1992-03-11
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 110
; LENGTH: 1247
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1247)
; US-09-177-776-110
```

```
; EARLIER FILING DATE: 1992-03-11
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 110
; LENGTH: 1247
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1247)
; US-09-178-115-110

Query Match
Best Local Similarity 4.9%; Score 34.4; DB 4; Length 1247;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 229 gacaggaacacccgcccctgtggaggagggtgagggccaccagcctgtgcacaaacaaca 288
DB 1100 GACTGTACCATTCGACTCTGGCCTGGGTGACAGCAAGACCCCTGTCTCAAAAAAGAAA 1041

QY 289 acaacaacacacacataaccacacacacacacacacacacacacacacacacacacac 328
DB 1040 ACAAAAACAAAAACAAAAACGCTACCTGCCCTGAGCAGG 1001

RESULT 13
US-09-177-776-110/c
; Sequence 110, Application US/0917776A
; Patent No. 6297051
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; FILE REFERENCE: D-0021.5A
; CURRENT APPLICATION NUMBER: US/09/177,776A
; EARLIER FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 08/787,739
; EARLIER FILING DATE: 1997-01-24
; EARLIER APPLICATION NUMBER: 08/485,049
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/486,756
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/477,504
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/481,658
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485,862
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485,863
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/487,077
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/260,190
; EARLIER FILING DATE: 1994-06-15
; EARLIER APPLICATION NUMBER: 08/177,093
; EARLIER FILING DATE: 1993-12-30
; EARLIER APPLICATION NUMBER: 07/964,589
; EARLIER FILING DATE: 1992-10-21
; EARLIER APPLICATION NUMBER: PV-709-92
; EARLIER FILING DATE: 1992-03-11
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 110
; LENGTH: 1247
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1247)
; US-09-177-776-110
```

CURRENT APPLICATION DATA:



```

?
? NUMBER OF SEQ ID NOS: 33
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 17
? LENGTH: 289
? TYPE: RNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Translation template
?
```

```
Query Match      8.5%; Score 59.8; DB 1; Length 7218;
Best Local Similarity 5.3%; Pred. No. 3.4e-08;
Matches 22; Conservative 227; Mismatches 164; Indels 0; Gaps 0;

QY 56 gccagcgagcgagcgagcattgactgaaacgacagagtcacgagcgagcgagcgaattg 115
Db 1467 GCAAGTAGTTAAGACATAGACAATTCGTACRRRRRRRRRRRRRRRRRRRRRRRRRR 1408

QY 116 acagaggactacgcccgaaggatagcccgagaaatagcaacacgagaaacaaatctcat 175
Db 1407 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1348

QY 176 catgccaattccaccaggaacgaagagctggagcagccctgcagatgagacagga 235
Db 1347 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1288

QY 236 acacgcccctgagagagtgagggccaccagcctgctgcaaacacacacacaa 295
Db 1287 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1228

QY 296 ccacacataacacacacacacagagccagggcctgcgcacccctgaaactt 355
Db 1227 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1168

QY 356 ccgatggccattcccaacagcgagatgaatgacgggttggtggagatggagatgat 415
Db 1167 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1108

QY 416 gaaatgttcaggaggagatagagagatccggagaaagcttagggagctac 468
Db 1107 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1055

RESULT 2
US-08-155-888-1
; Sequence 1, Application US/08155888
; Patent No. 6066623
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Hedstrom, Richard C.
; APPLICANT: Sedegah, Martha
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE PROTECTIVE
; TITLE OF INVENTION: AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR
; TITLE OF INVENTION: DELIVERING POLYNUCLEOTIDE VACCINES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Naval Medical Res. & Dev. Cmd.
; STREET: Bldg. 1, T-12 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/155.888
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, A. David
; REGISTRATION NUMBER: 24,743
; REFERENCE/DOCKET NUMBER: N.C. 75,851
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 295-6759
; TELEFAX: (202) 295-1022
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
; TOPOLOGY: circular
; HYPOTHEtical: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..755
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "promoter"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /label= CMV-IE
; OTHER INFORMATION: /note= "This feature acts as a promoter for any
; OTHER INFORMATION: downstream DNA sequence."
; OTHER INFORMATION: /citation= ([2])
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 933..2367
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 933
; OTHER INFORMATION: /function= "protein protective against malaria"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /number= 1
; OTHER INFORMATION: /label= IL2-CSP
; OTHER INFORMATION: /citation= ([1])
; PUBLICATION INFORMATION:
; AUTHORS: Sedegah, Martha
; AUTHORS: Hoffman, Stephen L.
; TITLE: Vaccination with Plasmodium yoelii CS protein
; TITLE: plasmid DNA protects against malaria
; JOURNAL: Science
; PUBLICATION INFORMATION:
; AUTHORS: Cullen, Bryan R.
; TITLE: TRANS-ACTIVATION OF HUMAN IMMUNODEFICIENCY
; TITLE: VIRUS OCCURS VIA A BIMODAL MECHANISM
; JOURNAL: CELL
; VOLUME: 46
; DATE: 26 SEP-1986
; PAGES: 973-982
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 4732
US-08-155-888-1

Query Match      6.4%; Score 44.6; DB 3; Length 5552;
Best Local Similarity 49.5%; Pred. No. 0.0013;
Matches 143; Conservative 0; Mismatches 144; Indels 2; Gaps 1;

QY 97 caggcgagcgagcgagcattgacagggaggtacgcccgaaggatagggcccaagaaatagcaa 156
Db 1770 CAAGGCCAGGAGCACCACAGGGGCCAGGAGCACCACCAAGGCGCAGGAGCACCACAAGCA 1829

QY 157 ccagga--aacaaaatctcatcatgccaatgtccaccaggaagaaacgaagagctggagca 214
Db 1830 CCAGGAGCACCACAGGGTCCAGGAGCACCACCAAGGAGCAGGAGCACCACCAAGGAGCAGGA 1889

QY 215 gccctctgagaatggacaggaacacgcctgtggaggaggtgagggccaccagcctgc 274
Db 1890 GCACCACAGGTCCAGGAGCACCACAGGGGCCAGGAGCAGGAGCACCACCAAGGAGCAGGAGCA 1949

QY 275 tgcaaacacacacacacacacacacacacacacacacacacacacacacacacacacacac 334
Db 1950 CAAGAACCACCCACACACACCCACCCACCAACCAACCAACCAACCAACCAACCAACCAACCA 2009

QY 335 tcgacgacctgcccctaaacttcgatgggccatttcccaacagggcagatg 383
Db 2010 CAACAGCCACCAACAGCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAATG 2058

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2002, 15:36:56 ; Search time 61.41 seconds  
(without alignments)  
2581.576 Million cell updates/sec

Title: US-09-327-750D-28

Perfect score: 700

Sequence: 1 acgagcgtctgcgcagcagc.....ggggcttgtgtccagtgga 700

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
c 1	59.8	8.5	7218	1	US-08-232-463-14, Appl
2	44.6	6.4	5552	3	US-08-155-888-1, Appl
3	38.4	5.5	7218	1	US-08-232-463-14, Appl
4	36.8	5.3	289	4	US-09-007-005-17, Appl
5	36.8	5.3	289	4	US-09-244-796-17, Appl
6	34.8	5.0	435	4	US-08-692-922-3, Appl
7	34.8	5.0	5163	3	US-08-700-651-1, Appl
8	34.8	5.0	5163	3	US-08-928-361B-4, Appl
9	34.8	5.0	5318	3	US-08-700-651-2, Appl
10	34.8	5.0	5318	3	US-08-928-361B-3, Appl
c 11	34.8	5.0	5917	4	US-08-692-922-1, Appl
c 12	34.4	4.9	1247	4	US-09-178-115-110, App
c 13	34.4	4.9	1247	4	US-09-177-776-110, App
c 14	34.4	4.9	2501	3	US-08-787-739-58, Appl
c 15	34.4	4.9	2501	4	US-09-178-115-58, Appl
c 16	34.4	4.9	2501	4	US-09-177-776-58, Appl
c 17	34.4	4.9	3532	3	US-08-787-739-90, Appl
c 18	34.4	4.9	3532	4	US-09-178-115-90, Appl
c 19	34.4	4.9	3532	4	US-09-177-776-90, Appl
c 20	34.4	4.9	10898	2	US-08-481-658B-5, Appl
c 21	34.4	4.9	10898	2	US-08-477-504A-5, Appl
c 22	34.4	4.9	10898	2	US-08-486-756A-5, Appl
c 23	34.4	4.9	10898	2	US-08-485-862B-5, Appl
c 24	34.4	4.9	10898	3	US-08-787-739A-5, Appl
c 25	34.4	4.9	10898	3	US-08-487-077A-5, Appl
c 26	34.4	4.9	10898	3	US-08-485-863A-5, Appl
c 27	34.4	4.9	10898	4	US-08-485-049D-5, Appl

c 28	34.4	4.9	10898	4	US-09-178-115-5, Appl
c 29	34.4	4.9	10898	4	US-09-177-776-5, Appl
c 30	34.2	4.9	449	4	US-08-688-908-3, Appl
31	34	4.9	198	5	PCT-US95-10668-1, Appl
32	34	4.9	198	5	PCT-US95-10668-2, Appl
33	34	4.9	198	5	PCT-US95-10668-3, Appl
34	34	4.9	198	5	PCT-US95-10668-4, Appl
35	33.6	4.8	3066	4	US-09-086-912-1, Appl
36	33.4	4.8	5511	3	US-08-928-361B-2, Appl
37	33.4	4.8	7334	3	US-08-928-361B-1, Appl
38	33.2	4.7	413	1	US-08-303-266-1, Appl
39	33.2	4.7	413	3	US-08-862-881-1, Appl
40	33.2	4.7	688	4	US-08-998-416-915, Appl
c 41	33.2	4.7	4279	4	US-09-041-886-22, Appl
42	33	4.7	2251	2	US-08-836-443-2, Appl
43	33	4.7	2648	2	US-08-836-443-1, Appl
c 44	32.8	4.7	475	2	US-08-623-906A-20, Appl
45	32.8	4.7	495	5	PCT-US95-11199-1, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935.313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 3.8%; Score 34; DB 2; Length 2277;  
Best Local Similarity 29.0%; Pred. No. 2.8;  
Matches 54; Conservative 38; Mismatches 94; Indels 0; Gaps 0;  
QY 572 gaattgtctgcgtatcccttatgggggagctctcttaatacaccatgaccatcatgatgaatt 631  
DB 2224 RNARYTGTGTTRTAYTGTTCNACNARYTGRTRAAACNKGTYTCNGTYTTRTTNCCYT 2165  
QY 632 ttgccttatgccttgactcctgccatttatcatgatgattaaactactgtgattcccgctgtt 691  
DB 2164 TNGCYTTTYTNCKNSWNACYTGTTCNSWNSWARYTGTGTTCYTCYTCYTCCTCCATGRT 2105  
QY 692 ttcttttcttgcatcttcctaatatgcctttactgacccgttttgctgtgaaccctatg 751  
DB 2104 TDATYTGNGGYTTNGGYTYYTTTNGGRTGNACNGGYTTNACYTTNCCYTTTTCNCKNARNC 2045  
QY 752 ttattt 757  
DB 2044 KDATYT 2039

Search completed: March 11, 2002, 16:34:16  
Job time: 3440 sec





	Matches	66;	Conservative	0;	Mismatches	49;	Indels	0;	Gaps	0;
Qy	242	ttcagttctctccttgccttggtctacttgttctcaagaataaacacaccagaaaaaaa	301							
Db	580	TTTTTTT TTTTTTT TTTTTTT TTTTTTT TTTGAAAAAAACCAACCAACAAG	521							
Qy	302	aaatctcatcgccaatatccaccaggaaaaacgaagagatgtagcagccta	356							
Db	520	ACATCTCTCAATAAAAAAAAAAAGCATATAAAAAAAAAGGAACGCCGAAGACGGTG	466							

```

RESULT 11
US-08-993-674A-22/c
; Sequence 22, Application US/08993674A
; Patent No. 6228372
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,674A
; FILING DATE: 18-DEC-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-993-674A-22

```

	Query Match	4.1%	Score 36.6	DB 4	Length 581
	Best Local Similarity	57.4%	Pred. No. 0.23		
	Matches 66	Conservative 0	Mismatches 49	Indels 0	Gaps 0
QY	242	ttcagttctctctccttgcctttgtctctacttgtgttcaagagaaacaacacagaaaaa	301		
Db	580	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAAAAAACAACAACAACAG	521		
QY	302	aaatctcatcatgycaaatattccaccaggaaaaacgaagagatgagcagcctatg	356		
Db	520	ACATCTTCTCAATAAAAAAAGACATATAAAAAAAGAACGCCGAAGACGGTG	466		

RESULT 12  
US-08-332-766A-42  
; Sequence 42, Application US/08332766A  
; Patent No. 5843647

GENERAL INFORMATION:  
APPLICANT: JEFFREYS, Alec J.  
APPLICANT: ARMOUR, John  
TITLE OF INVENTION: SIMPLE TANDEM REPEATS  
NUMBER OF SEQUENCES: 125  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/332,766A  
FILING DATE: 01-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9326052.9  
FILING DATE: 21-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BIRD, Donald J.  
REGISTRATION NUMBER: 25,323  
REFERENCE/DOCKET NUMBER: 217211/M94/0434/GB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 278 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-332-766A-42

	Query Match	4.0%	Score 35.8;	DB 2;	Length 278;
	Best Local Similarity	61.1%;	Pred. No. 0.26;		
	Matches 58;	Conservative 0;	Mismatches 37;	Indels 0;	Gaps 0;
QY	55	ggaagaatccccgggaacgaaaaatggtggtttgggggaaggaggtaaaggagagaaa	114		
Db	158	GGAAGAAGAGACGGGAAAGAAAGAGGGAGGGAGGAGGGAGGGAGGAGGAGG	217		
QY	115	gctgagaggaggggccttaattggagcccccgtag	149		
Db	218	GAGGAGGGAGGGAGTATTAAGATGATCCCTTAG	252		

```

RESULT 13
US-08-676-967-2/c
Sequence 2, Application US/08676967
Patent NO. 5747317
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```





TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 102:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40328 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-742-185-102

Query Match 5.0%; Score 44.4; DB 3; Length 40328;  
Best Local Similarity 51.0%; Pred. No. 0.012;  
Matches 105; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 684 cgcgtgtttcttcttcctgcaatttctcaatgcctttactgacgtttgtgtgta 743  
DB 20178 CCCGCGCCCTTTTGTTCATTGCTGTGGTAGATCTTCCTGCATCCTTTTATTGGA 20237  
QY 744 accctatgtattccatgtgtcaagtgggtctgtgttgccagcttctattgaagatt 803  
DB 20238 GCCTATGTGTCTCTGACGTGAGATGGGTTTCTCGAATACAGCACACTGATGGGTCTT 20297  
QY 804 gctttgcactcagtgtaatttctgtcagcagtagtttccaccatttgcattggaataat 863  
DB 20298 GACTCTTATTCAGTTGCCAGCTGTGTCTTTTAATTGGAGCATTAGTCCATTGCAT 20357  
QY 864 ttaagccaataaagaacatttaaaaa 889  
DB 20358 TTTAAAGTTAATATTGTATGTATGAA 20383

RESULT 4  
US-08-416-478A-1/c  
; Sequence 1, Application US/08416478A  
; Patent No. 5773578  
; GENERAL INFORMATION:  
; APPLICANT: Hercend, Thierry  
; APPLICANT: Triebel, Frederic  
; TITLE OF INVENTION: New Proteins Produced By Human  
; TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And  
; TITLE OF INVENTION: Their Pharmaceutical And Biological Uses  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.  
; STREET: 419 Seventh Street, N.W., Suite 400  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/416.478A  
; FILING DATE: 04-APR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/854,644  
; FILING DATE: 08-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: YUN, Allen C.  
; REGISTRATION NUMBER: 37,971  
; REFERENCE/DOCKET NUMBER: HERCEND-1A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-5197  
; TELEFAX: (202) 737-3528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1871 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 231..1724  
US-08-416-478A-1  
  
Query Match 4.2%; Score 37.2; DB 1; Length 1871;  
Best Local Similarity 57.9%; Pred. No. 0.29;  
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;  
QY 14 ctccataccggtcctccattttgtgctgcaagctctgggaagaatcccggaac 73  
DB 239 CTCCACATCTCTCTATGGTCTGGCAGCGCTGAGCCCTCCAAAAGGAGGTCAGAAAG 180  
QY 74 gaaaaatggtggtttgggggaaggaggtgaaggaggagaaagctggagggagg 127  
DB 179 GAAAGGGGCACTGGTGGGGGTGGGGGTAAAGGAGAGATTCTCCAGAGAGGG 126  
  
RESULT 5  
US-08-474-988B-1/c  
; Sequence 1, Application US/08474988B  
; Patent No. 5874250  
; GENERAL INFORMATION:  
; APPLICANT: Hercend, Thierry  
; APPLICANT: Triebel, Frederic  
; TITLE OF INVENTION: NEW PROTEINS PRODUCED BY HUMAN  
; TITLE OF INVENTION: LYMPHOCYTES, DNA SEQUENCES ENCODING THESE PROTEINS AND  
; TITLE OF INVENTION: THEIR PHARMACEUTICAL AND BIOLOGICAL USES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.  
; STREET: 419 Seventh Street, N.W., Suite 400  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/474.988B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/416.478  
; FILING DATE: 04-APR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/854,644  
; FILING DATE: 08-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: YUN, Allen C.  
; REGISTRATION NUMBER: 37,971  
; REFERENCE/DOCKET NUMBER: HERCEND=1B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-5197  
; TELEFAX: (202) 737-3528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1871 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 231..1724



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2002, 16:33:56 ; Search time 61.41 seconds  
(without alignments)  
3285.978 Million cell updates/sec

Title: US-09-327-750D-29  
Perfect score: 891  
Sequence: 1 accccatccccactctat.....aataagcaatttataaaagc 891

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues  
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTBUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	62.4	7.0	7218	1	US-08-232-463-14
C 2	47.8	5.4	7218	1	US-08-232-463-14
C 3	44.4	5.0	40328	3	US-08-742-185-102
C 4	37.2	4.2	1871	1	US-08-416-478A-1
C 5	37.2	4.2	1871	2	US-08-474-988B-1
C 6	37.2	4.2	1871	2	US-08-394-442B-1
C 7	37	4.2	289	4	US-09-007-005-17
C 8	37	4.2	289	4	US-09-244-796-17
C 9	36.6	4.1	581	2	US-08-557-309B-22
C 10	36.6	4.1	581	3	US-08-834-306-22
C 11	36.6	4.1	581	4	US-08-993-674A-22
C 12	35.8	4.0	278	2	US-08-332-766A-42
C 13	34	3.8	2277	1	US-08-676-967-2
C 14	34	3.8	2277	1	US-08-676-974-2
C 15	34	3.8	2277	2	US-09-098-487-2
C 16	33	3.7	53526	3	US-08-658-136-2
C 17	33	3.7	53577	3	US-08-658-136-1
C 18	32.2	3.6	4517	4	US-09-140-804-9
C 19	32.2	3.6	4731	3	US-08-488-706-2
C 20	32.2	3.6	4731	3	US-08-772-270A-9
C 21	32.2	3.6	5852	1	US-07-867-106-2
C 22	32	3.6	289	4	US-09-007-005-17
C 23	32	3.6	289	4	US-09-244-796-17
C 24	32	3.6	623	4	US-09-385-982-167
C 25	32	3.6	246240	2	US-08-724-394A-20
C 26	32	3.6	246240	2	US-08-724-394A-21
C 27	32	3.6	246240	2	US-08-724-394A-22

28	31.8	3.6	772	2	US-08-889-337-2	Sequence 2, Appli
29	31.8	3.6	13146	2	US-08-724-354D-3	Sequence 3, Appli
30	31.8	3.6	13146	3	US-09-270-984A-3	Sequence 3, Appli
C 31	31.6	3.5	24979	2	US-08-147-777-3	Sequence 3, Appli
C 32	31.6	3.5	24979	3	US-08-452-872-3	Sequence 3, Appli
C 33	31.6	3.5	24979	5	PCT-US93-03985-3	Sequence 3, Appli
C 34	31.4	3.5	990	3	US-08-921-209-3	Sequence 3, Appli
C 35	31.4	3.5	990	3	US-09-411-763-3	Sequence 3, Appli
C 36	31.4	3.5	1000	2	US-08-827-613-1	Sequence 1, Appli
C 37	31.2	3.5	3695	4	US-09-211-704A-1	Sequence 1, Appli
C 38	31.2	3.5	35060	3	US-08-814-095-7	Sequence 7, Appli
C 39	31	3.5	983	3	US-09-048-889-6	Sequence 6, Appli
C 40	31	3.5	13146	2	US-08-724-354D-3	Sequence 3, Appli
C 41	31	3.5	13146	3	US-09-270-984A-3	Sequence 3, Appli
C 42	31	3.5	87350	3	US-08-781-891-79	Sequence 79, Appl
C 43	30.8	3.5	3396	4	US-08-974-549A-640	Sequence 640, App
C 44	30.8	3.5	6678	3	US-08-816-617A-1	Sequence 1, Appli
C 45	30.6	3.4	32207	2	US-08-770-379-20	Sequence 20, Appli

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
: Sequence 14, Application US/08232463  
: Patent No. 5670367  
: GENERAL INFORMATION:  
: APPLICANT: DORNER, F.  
: APPLICANT: SCHEIFLINGER, F.  
: APPLICANT: FALKNER, F. G.  
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
: NUMBER OF SEQUENCES: 52  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Foley & Lardner  
: STREET: 1800 Diagonal Road, Suite 500  
: CITY: Alexandria  
: STATE: VA  
: COUNTRY: USA  
: ZIP: 22313-0299  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/232,463  
: FILING DATE:  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US/07/935,313  
: FILING DATE:  
: APPLICATION NUMBER: EP 91 114 300.6  
: FILING DATE: 26-AUG-1991  
: ATTORNEY/AGENT INFORMATION:  
: NAME: BENT, Stephen A.  
: REGISTRATION NUMBER: 29,768  
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (703)836-9300  
: TELEFAX: (703)683-4109  
: TELEX: 899149  
: INFORMATION FOR SEQ ID NO: 14:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 7218 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: IMMEDIATE SOURCE:  
: CLONE: pTZgpt-Fls  
US-08-232-463-14

**THIS PAGE BLANK (USPTO)**

Qy 806 ctttgcaactcagtgtaagtttctgtcagcagtagtttccaccatttgcattgaggaataattt 865  
 Db 86 CTTTGCACTCAGTGTAAAGTTTCTGTGACGAGTAGTTTCCACCCATTGTGCATGGAAAAATTT 27  
 Qy 866 aaagccaataagaattttataaagc 891  
 Db 26 AAAGCTAATAAAGCAATTTAAAAAGC 1

RESULT 15  
 BF732486/c  
 LOCUS  
 DEFINITION  
 naelld11.x1 NCI-CGAP\_Ov18 Homo sapiens cDNA clone IMAGE:3434948 3',  
 similar to SW:HG74\_HUMAN O00994 OVARIAN GRANULOSA CELL 13.0 KD  
 PROTEIN HG74. ;, mRNA sequence.

ACCESSION  
 BF732486  
 VERSION  
 BF732486.1 GI:12057561  
 KEYWORDS  
 EST.  
 SOURCE  
 human.

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 566)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento  
 Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The  
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
 Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 info@image.llnl.gov  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 453.

# FEATURES

source  
 1..566  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3434948"  
 /clone\_lib="NCI-CGAP\_Ov18"  
 /tissue\_type="fibrotheoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site:1: Not 1; Site:2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTTACCAATCTGAAGTGGAGCGCGCGGACATTTTTTTTTTTTTTTT 3'];  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not  
 I and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 162 a 140 c 110 g 153 t 1 others

# BASE COUNT

ORIGIN  
 Query Match 62.9%; Score 560.8; DB 11; Length 566;  
 Best Local Similarity 99.5%; Pred. No. 1.9e-106;  
 Matches 562; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 327 caggaaaacgaagatggagcagcctatgcagaatgagagagacgccgcctttggga 386  
 Db 566 CAGGAAACGAAGACAGATGCGAGCCTATGCAGATGAGAGAGACCGCCCTTTGGGA 507  
 Qy 387 ggaggtgaagccaccagcctgcaggaataacaggggacagggctcgcgacttgcacct 446  
 Db 506 GAAGGTGAAGGCCACCAGCCTGCAGGAATCGACGGGACAGGCTCGCGCGACTTGCCTCT 447

Search completed: March 11, 2002, 16:09:59  
 Job time: 10894 sec

Qy 447 aatttccgatgggccataccccaataggcagatcaatgatgggatgggtggagatggagat 506  
 Db 446 AATTTTCGATGGCCATACCCAATAGGCAGATCAATGATGGGATGGGTGGAGATGAGAT 387  
 Qy 507 gatatggaaaattatcatggagagatgagagaaaatcagaagaaaacttgggagctgcag 566  
 Db 386 GATATGGAATATTTCATGGAGGAGATGAGAGAAATCAGAAGAAACTTAGGGAGCTGCAG 327  
 Qy 567 ttgaggaattgtcgtatcctctatggggagcctctctaatcaccatgaccatcatgat 626  
 Db 326 TTGAGGAATTGTCTGCGTATCCTTATGGGGAGGCTCTCTAATCACCATGACCATCATGAT 267  
 Qy 627 gaatttgccttatgccttgactcctgccttattatcatgatgagataactgtgattcccg 686  
 Db 266 GAATTTTGCTTATGCCCTTGACTCCTGCCATTTATCATGAGATTAATACTGTGATTCGCG 207  
 Qy 687 ctggtttcttcttcccttgcatatttctctaatatgccttactgacogtttgcgtgaaacc 746  
 Db 206 CTGTTTCTCTTTTCCCTTGCATTTTCTTAATATGCCCTTTACTGATCGTTTGTGTGAACC 147  
 Qy 747 ctatgttatttccatgtgtcaagtggtctctgttgcagcttctctatttgaagattgcc 806  
 Db 146 CTATGTATTATTCATGTGTCAAGTGGGCTCTGTGTGGCAGCTTCTATTTGAAGATTGCC 87  
 Qy 807 ttgcactcagtgtaagtttctgtcagcagtagtttccaccatttgcattgggaaaaattta 866  
 Db 86 TTTGCACTCAGTGTAAAGTTTCTGTGACGACGATAGTTTACCCCATTTGCATGGAAAAATTTA 27  
 Qy 867 aagccaataaagcaattttaaaagc 891  
 Db 26 AAGCTAATAAAGCAATTTAAAAAGC 2







cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Insert Length: 759 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 451.

# FEATURES

```

1. 574
Location/Qualifiers
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:1640433"
  /clone_lib="Soares_testis_NHT"
  /sex="male"
  /lab_host="DH10B"
  /note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 162 a 141 c 111 g 160 t
ORIGIN
Query Match 63.3%; Score 564.4; DB 10; Length 574;
Best Local Similarity 99.8%; Pred. No. 3.4e-107;
Matches 565; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 326 ccaggaacacgaagagatggagcagcctatgcagaatggaggaagacgcgcctttggg 385
|||||
Db 574 CCAGAAAACGAAGAGATGGAGCAGCCTATGCAGAAATGGAGAGAGACGCCCTTTGGG 515
|||||

QY 386 aggaagtgaagccaccagcctgcaggaatcgacggagacagctcgccgacttgcgc 445
|||||
Db 514 AGAGGTGAAGCCACACAGCTGCAGGAATCGACGGGACAGGCTCGCGACTTGCCTCC 455
|||||

QY 446 taatttcgatgggcatcaccatagcgagatcaatgatggatgggtggagatggaga 505
|||||
Db 454 TAATTTTCATGGCGCATACCCAAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGA 395
|||||

QY 506 tgatatggaatatcatcattgagagatgagagaatcagagaacaaacttagggagctgca 565
|||||
Db 394 TGATATGGAATATTCATGTGGAGGAGATGAGAGAAATCAGAGAAACTTAGGGAGCTGCA 335
|||||

QY 566 gttgaggaattgtctgcgtatccttatggggagctctctaatcaccatgaccatcatga 625
|||||
Db 334 GTTGAGGAATTCCTGCGTATCCTTATGGGGAGGCTCTTAATCACCATGACCATCATGA 275
|||||

QY 626 tgaatttgccttactgccttgccttacttgccttactgccttactgccttactgccttcc 685
|||||
Db 274 TGAATTTTGCCTTATGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTT 215
|||||

QY 686 gctgtttctttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcc 745
|||||
Db 214 GCTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 155
|||||

QY 746 cctatgtatttccatgtgtcagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 805
|||||
Db 154 CCTATGTTATTTCCATGTGTCAAGTGGGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 95
|||||

QY 806 ctttgcactcagtgtaagtcttctcagcagtagtttccaccatttgcattgaggaataatt 865
|||||
Db 94 CTTTGACATCAGTGAAGTTCTGTGCAGCAGTAGTTTCCACCATTTTGCATGGAATAATTT 35
|||||

QY 866 aaagccaataaagcaatttaaaagc 891
|||||
Db 34 AAAGTAATAAAGCAATTTAAAAAGC 9
|||||

```

## RESULT 11

```

AV709328 610 bp mRNA EST 09-OCT-2000
LOCUS AV709328 ADC Homo sapiens cDNA clone ADCALB10 5', mRNA sequence.
DEFINITION AV709328
ACCESSION AV709328
VERSION AV709328.1 GI:10726849
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 610)
AUTHORS Peng, Y., Song, H., Huang, Q., Huang, C., Gu, Y., Yang, Y., Gao, G., Xiao
, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z.,
Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu
, G., Hu, R., Chen, J., Chen, Z. and Han, Z.
TITLE Homo sapiens cDNA ADC clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@hgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..610
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ADCALB10"
/clone_lib="ADC"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 181 a 115 c 148 g 166 t
ORIGIN
Query Match 63.3%; Score 564.4; DB 10; Length 610;
Best Local Similarity 99.8%; Pred. No. 3.4e-107;
Matches 565; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 326 ccaggaacacgaagagatggagcagcctatgcagaatggaggaagacgcgcctttggg 385
|||||
Db 22 CCAGAAAACGAAGAGATGGAGCAGCCTATGCAGAAATGGAGAGAGACGCCCTTTGGG 81
|||||

QY 386 aggaagtgaagccaccagcctgcaggaatcgacggagacagctcgccgacttgcgc 445
|||||
Db 82 AGAGGTGAAGCCACACAGCCTGCAGGAATCGACGGGACAGGCTCGCGACTTGCCTCC 141
|||||

QY 446 taatttcgatgggcatcaccatagcgagatcaatgatggatgggtggagatggaga 505
|||||
Db 142 TAATTTTCATGGCGCATACCCAAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGA 201
|||||

QY 506 tgatatggaatatcattgagagatgagagaatcagagaacaaacttagggagctgca 565
|||||
Db 202 TGATATGGAATATTCATGTGGAGGAGATGAGAGAAATCAGAGAAACTTAGGGAGCTGCA 261
|||||

QY 566 gttgaggaattgtctgcgtatccttatggggagagctctctaatcaccatgaccatcatga 625
|||||
Db 262 GTTGAGGAATTCCTGCGTATCCTTATGGGGAGGCTCTTAATCACCATGACCATCATGA 321
|||||

QY 626 tgaatttgccttactgccttgccttgccttgccttactgccttactgccttactgccttcc 685
|||||
Db 322 TGAATTTTGCCTTATGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTT 381
|||||

QY 686 gctgtttctttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcc 745
|||||

```



Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL: contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 437.

## FEATURES

source  
1. .662  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2519611"  
/clone\_lib="Schneider fetal brain 00004"  
/sex="male"  
/tissue\_type="frontal lobe"  
/dev\_stage="5 months post-conception"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescript SK (Stratagene);  
Site\_1: SstI; Site\_2: XhoI; Double-stranded cDNA was  
prepared from human fetal brain tissue. 5' and 3'  
adaptors were used in cloning as follows: 5' adaptor  
sequence:  
5'-GAGACAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCGCCCCCCC-3'  
and 3' adaptor sequence:  
5'-GAGACAGAGAGCTGAGCTTTT-3'. The library was  
size-selected for >0.5 kb inserts and has an average  
insert size estimated at 1.2 kb. This library was  
constructed using the CAP-trapper method for full-length  
enrichment and has not undergone amplification. Library  
was constructed by Dr. Claudio Schneider (LNCIB-Area  
Science Park, Trieste, Italy). 2 others  
178 a 143 c 169 g 170 t

## BASE COUNT

178 a 143 c 169 g 170 t

Query Match 64.4%; Score 574; DB 10; Length 662;

Best Local Similarity 98.8%; Pred. No. 3.4e-109; Mismatches 7; Indels 0; Gaps 0;

Matches 577; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 281 agaaaaaacacagaaaaaaatctcatcgcaaatattccaccaggaacgaaga 340

Db 65 AGAAAAACAACAGAAAAAATCTCATCGCAATATTCACCGAGAAACGAGCA 124

QY 341 gatgagagcctatgcagaaatggagaggaacccctttggaggagggaagccca 400

Db 125 GATGGAGCAGCCTATGCAGAAATGGAGGAGAACCCGCTTTGGGAGGAGTGAAGGCCA 184

QY 401 ccagcctcagaaatcagcgggagcagctcgcgaacttgccttaatttcgatgggc 460

Db 185 CCAGCCTGCAGGAATCGACGGGGACAGCTCGCGACTTGCCTTAATTTTCGATGGGC 244

QY 461 cataccataggcagatcaatgatggatgggtggagatggagatgatggaataatt 520

Db 245 CATACCCATAGGCAGATCAATGATGGGATGGTGGAGATGGAGATGATGGAATATT 304

QY 521 catggaggagatgagagaatcagaagaaacttagggagctgcagttgaggaattgtct 580

Db 305 CATGGAGGAGATCAGAGAAATCAGAAGAAACTTAGGGAGCTGCAGTTCAGGAATTGTCT 364

QY 581 gcgtatccttatggggagcctcttaatacatgacatcatgatgaatttgccttat 640

Db 365 GCGTATCTTATGGGGAGCTCTTAATCACCATGACCATCATGATGAATTTGCTTAT 424

QY 641 gcttgactcctgcattatcatgatgattaactgtgattccgcgtgtttcttttc 700

Db 425 GCGTTGACTCTCGCCATTTATCATGAGATTATAGTGTATTCGCCCTGTTTCTTNTC 484

QY 701 cttgcaatttcctaataatgccttactgacccggtttgctgtgaacctatgtattcca 760

Db 485 CTTGCAATTCCTAATATGCGTTTACTGATCGGTTGCTGTGAACCTATGTTATTTCCA 544

QY 761 tgtgtcaagtggcttctgtgtgcagctctctatttgaagattgcctttgcaactcagtt 820

Db 545 TGTGTCAAGTGGGCTCTGTGTGCCAGCTTCATTTTGAAGATTGCGCTNTGCTCAGTGT 604

QY 821 aagtttcttcagcagtagtttcaccatttgcattgcatggaataatt 864

Db 605 AAGTTCCTCTCAGCACTAGTTTCACCCATCTGCGATGGAACATTT 648

## RESULT

AA514324/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

603 bp mRNA EST 18-AUG-1997  
nf56b04.s1 NCI\_CGAP\_Co3 Homo sapiens cDNA clone IMAGE:923887 3'  
similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74  
(HUMAN);, mRNA sequence.

AA514324  
AA514324.1 GI:2253832  
EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 603)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.  
Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/dbrrp/image/image.html](http://www-bio.llnl.gov/dbrrp/image/image.html)  
Insert Length: 840 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 433.

## FEATURES

source

1. .603

/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="IMAGE:923887"

/clone\_lib="NCI\_CGAP\_Co3"

/sex="pooled"

/tissue\_type="colon"

/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from 12 pooled bulk tumor samples and primed  
with a Not I - Oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT73 vector. Library went through one round of  
normalization."

## BASE COUNT

166 a 141 c 116 g 180 t

Query Match 63.4%; Score 565.2; DB 10; Length 603;

Best Local Similarity 99.3%; Pred. No. 2.3e-107;

Matches 578; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 311 catggcaaatattccaccagg-aaaacgaagatgagcagcctatgcagaatggagagg 369

Db 603 CATGGCAATATATTCACGAGAAACGAGATGGAGACCCTATGCAAGATGGAGG 544

QY 370 aagaccgcctttggaggagggtgaaggccaccagcctgcaggaatcgacggggacagg 429

Db 543 AAGACCGCCTTTGGGAGGAGGTGAAGCCACCAGCCTGCAGGAATCGACGGGACAG 484

QY 430 ctgcgcgaacttgccttaatttttcgatggccatacccaatagggcagatcaatgatggga 489

Db 483 CTCGCCGACTTGCCCTTAATTTTCGATGGGCCATACCCAATAGGCAGATCAATGATGGGA 424

QY 672 atactgtgattcccgctgttttcttcttcttgcatttccctaataatgctttactgac 731  
 |||||  
 Db 226 ATACTGTGATTCCCGCTGTTTCTTTCTTCTTGCATTTTCTTAATATGCTTTACTGATC 167  
 |||||  
 QY 732 cgttgcgtgaaacctatgttatttccatgctgtaagtgggtctctgtgttgcagcttc 791  
 |||||  
 Db 166 CGTTTGCTGTGAACCTATGTTATTTCCATGTCTCAAGTGGGTCTGTGTTGCCAGCTTC 107  
 |||||  
 QY 792 tattgaagattgccttgcactcagtgtaagtcttctctgcagcagtagtttcaaccattt 851  
 |||||  
 Db 106 TATTTGAAGATTGCTTTCACCTCAGTGTAAAGTTTCTGTGACAGTAGTTCACCCATTT 47  
 |||||  
 QY 852 gcatgaaaaatttaagccaataaagcaatttaaaaaagc 891  
 |||||  
 Db 46 GCATGAAAAATTTAAAGCTATATAAGCAATTTTAAAGC 7  
 |||||

RESULT 6  
 AI937016/c  
 LOCUS  
 DEFINITION  
 wp72b04.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2467279.3'  
 similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74  
 (HUMAN);, mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 606)  
 NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>  
 National Cancer Institute / National Institute of Neurological  
 Disorders and Stroke, Brain Tumor Genome Anatomy Project  
 (CGAP/BTGP), Tumor Gene Index  
 Unpublished (1998)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
 Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)  
 Insert Length: 783 Std Error: 0.00  
 Seq primer: -400p from Gibco  
 High quality sequence stop: 457.  
 Location/Qualifiers  
 1..606  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2467279"  
 /clone\_lib="NCI-CGAP\_Brn25"  
 /tissue\_type="anaplastic oligodendroglioma"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTACCAACTGTAAGTGGAGCGCGCATAGCTTTTCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M.Fatima Bonaldo."

BASE COUNT 168 a 144 c 119 g 173 t 2 others  
 ORIGIN

Query Match

64.8%; Score 577; DB 10; Length 606;

Best Local Similarity 97.9%; Pred. No. 8.2e-110;  
 Matches 594; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 285 aaacaacagaaaaaaatctcatcgcaataattaccaggaaacagaagatg 344  
 |||||  
 Db 606 AAACACACAAAAAAATCTCATATGGCAATAA-TCAACAGGANAACAGAGATG 548  
 |||||  
 QY 345 gagcagcttatcgaaatgagagaaagaccgcccttgggagaggtgaaggccaccag 404  
 |||||  
 Db 547 GAGCAGCCTATCGAATGGAGAGGAGACCCCTTTGGGAGGAGGTGAAGCCACAG 488  
 |||||  
 QY 405 cctgcagaaaaatcgacggggacaggtcgccgacttgccctaaatttctcgatgggcata 464  
 |||||  
 Db 487 CCTGCAGGAAATCGACGGGACAGGCTCGCCGACTTGCCCTAAATTTTCATGGGCCATA 428  
 |||||  
 QY 465 cccaatagcgagatcaatgatgggatgggtgagatggagatgataatattatcatg 524  
 |||||  
 Db 427 CCCAATAGGCAGATCAATGATGGGATGGGTGGAGATGGAGATCATATGGAATATTTCATG 368  
 |||||  
 QY 525 gagagatgagagaaatcagaagaaacttagggagctgagttggggaattgtctgcgt 584  
 |||||  
 Db 367 GAGGAGATGAGAGAAATCAGAAAGAAACCTTAGGAGCTGCAGTTGAGGAATTTGCTGCCCT 308  
 |||||  
 QY 585 atccttatggggagctctcttaatacaccatgaccatcatgatgaattttgccttatgcct 644  
 |||||  
 Db 307 ATCCTTATGGGGAGCTCTCTTAATCACCATGACCATCATGATCAATTTTGCCTTATGCCT 248  
 |||||  
 QY 645 tgactcctgcatttatcatgagatttaactatgattcccgctgtttcttcttcttctg 704  
 |||||  
 Db 247 TGACTCCTGCCAATTTATCATGAGATTAATPACTGTGATTCCCGCTGTGTTTTCTTTTCTTTG 188  
 |||||  
 QY 705 catttcttaataccttctactgacgttctgctgtaaacctatgttatttccatgctg 764  
 |||||  
 Db 187 CATTTTCTTAATATGCTTTTACTGATCCGTTTCTGTGAACCTATGTTATTTCATGATG 128  
 |||||  
 QY 765 tcaagtgggtctgtgtgcccagcttctatttgaagattgaccttgcactcagtgtaagt 824  
 |||||  
 Db 127 TCAAGTGGGTCTGTGTTGCCAGCTTCTATTGGAAGATTGCTTTGCACCTAGTGTAACT 68  
 |||||  
 QY 825 ttctctcagcagtagtttccaccatttgcattggagaaaaatttaagccaataagaattt 884  
 |||||  
 Db 67 TTCTCTCAGCAGTAGTTTTCACCCATTTGCATCGAAAAATTTAAAGCTAATAAAGCAATTT 8  
 |||||

QY 885 aaaaagc 891  
 |||||  
 Db 7 AAAAAAGC 1

RESULT 7  
 AI929106  
 LOCUS  
 DEFINITION  
 au5b10.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone  
 IMAGE:2519611.5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0  
 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 662)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin  
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
 White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project  
 Unpublished (1997)  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 339.  
Location/Qualifiers  
1. 620  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:107472"  
/clone\_lib="NCI\_CGAP\_Co9"  
/tissue\_type="colon tumor RER+"  
/lab\_host="DH10B"  
/note="organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from RER+ colon tumor, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo (Soares4).

FEATURES  
source

BASE COUNT 169 a 144 c 116 g 191 t  
ORIGIN

Query Match 66.9%; Score 596.4; DB 10; Length 620;  
Best Local Similarity 99.8%; Pred. No. 7.8e-114;  
Matches 597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 294 gaaaaaaatctctatcgcaaatattcaccaggaacgaagagatgagcagcct 353  
|||||  
Db 620 GAAAAAAAATCTCATCTGTCGCAATATTCACCGAGAAACGAGAGATGAGCAGCCT 561  
QY 354 atcgagaatggaggaagaccgcccctttgggagaggtgaaggccaccagcctcgagga 413  
|||||  
Db 560 ATGCAGATGGAGAGGAGAACCCCTTTGGGAGGAGGTGAAGGCCACCAGCCTGCAGGA 501  
QY 414 aatcgacggggacaggctcgagacttgccttaatttccgatgggccaatacccaatagg 473  
|||||  
Db 500 AATCGACGGGGACAGGCTCGCGACTTGCCTTAATTTTCGATGGGCCATACCCCAATAGG 441  
QY 474 cagataatgatggatgggtggagatgagatgatgatgaaatattctatcgagagatg 533  
|||||  
Db 440 CAGATCAATGATGGATGGGTGGAGATGAGATGATATGCAATATTCATGAGGAGATG 381  
QY 534 agagaatcagaagaaacttggagagctgcagttgaggaattgtctcgatcctctatg 593  
|||||  
Db 380 AGAGAAATCAGAAGAAACTTGGGAGCTGCAGTTGAGGAATTTGCTGCTATCCTTATG 321  
QY 594 ggggagctcttaaccaccatgaccatgatgaatttggccttatgcttgcacttgcactctg 653  
|||||  
Db 320 GGGGAGCTCTTAATCACCATGACCATCATGATGAATTTTGCTTATGCTTGCCTTGCCTCTG 261  
QY 654 ccattttatcagatttaatactgtgattcccgctgtttcttttcccttgcatcttct 713  
|||||  
Db 260 CCATTATCATGAGATTATGATGATTCCTGCTGCTTTCTTTTCCCTGCAATTTCTCT 201  
QY 714 atatgcctttactgatccgtttgtgtgaaccctatgttatttccatgtgcaagtggg 773  
|||||  
Db 200 AATATGCCCTTACTGATCGCTTGTGCTGAACCTATGTTATTTCCATGTGCAAGTGGG 141  
QY 774 tcttggttgcagcttctatttgaagattgcttggactcagtgatgaagttctctgctag 833  
|||||  
Db 140 TCTTGTGTTGCCAGCTTCCTATTGGAAGATTGCTTTGCACTCAGTCAAGTTTCTGCTAG 81  
QY 834 cagtagttccaccatttgcattggaaaaattttaaagccaataaagcaattttaaagc 891  
|||||  
Db 80 CAGTAGTTCCACCATTTCGATGGAAAAATTTAAAGCTAATAAAGCAATTTAAAAAGC 23

RESULT 5  
BF058872/c  
LOCUS  
DEFINITION  
BF058872 586 bp mRNA EST 16-OCT-2000  
7k35g02.x1 NCI\_CGAP\_Ov18 Homo sapiens cDNA clone IMAGE:3477674 3'  
similar to SW:HG74\_HUMAN Q00994 OVARIAN GRANULOSA CELL 13.0 KD

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

FEATURES  
source

PROTEIN HGR74. ; mRNA sequence.  
BF058872  
BF058872.1 GI:10812768  
EST  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 586)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@remail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento  
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
[info@image.llnl.gov](mailto:info@image.llnl.gov)  
Seq primer: -400P from Gibco  
High quality sequence stop: 464.  
Location/Qualifiers  
1. 586  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3477674"  
/clone\_lib="NCI\_CGAP\_Ov18"  
/tissue\_type="fibrotheoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="organ: ovary; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAATGGAGCGCCGCGACATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 166 a 143 c 112 g 164 t  
ORIGIN

Query Match 64.8%; Score 577.4; DB 11; Length 586;  
Best Local Similarity 99.7%; Pred. No. 6.8e-110;  
Matches 578; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 312 atggcaaatattcaccaggaacgaagagatgagcagcctatgcagaatggagagaa 371  
|||||  
Db 586 ATGGCAAAATATTCACCGAGAAACGAGAGATGAGCAGCAGTATGCGAATGGAGAGAA 527  
QY 372 gaccgcccctttgggagaggtgaagccaccagcctgcaggaaatgcaggggacaggct 431  
|||||  
Db 526 GACCGCCCTTTGGGAGGAGTGAAGCCACCAGCCTGCAGGAAATCGACGGGCACAGCT 467  
QY 432 cgccgacttgcctcctaatttccgatgggccaatacccaataggcagatcgaatgatg 491  
|||||  
Db 466 CCGCGACTTGCCTTAAATTTTCGATGGCCATACCCAAATAGGCAGATCAATGATGGGATG 407  
QY 492 ggtggagatggagatgatgaaatattcattcagagagatgagaaatcagaagaaa 551  
|||||  
Db 406 GTGGAGATGGAGATGATATGGAATATTCATGAGGAGATGAGAGAAATCAGAGAGAAA 347  
QY 552 cttagggagctcagttgaggaattgtctgctatctcttatggggagctctctaatcac 611  
|||||  
Db 346 CTTAGGGAGCTGCAGTTGAGGAATTTCTGCGTATCTTATGGGGAGCTCTCTAATCAC 287  
QY 612 catgaccatcatgatgaatttgccttatgcttgcactgcctgacttcattatcatgatgatta 671  
|||||  
Db 286 CATGACCATCATGATGAATTTTGCCTTATGCTTGCCTTGCCTTATCATGAGATTATTA 227





/lab\_host="DH10B"  
/note="Organ: lymph node; Vector: PCMV-SPORT6; Site\_1:  
Salt; Site\_2: NotI; Cloned unidirectionally. Primer:  
Oligo dT. Average insert size 1.25 kb. Life Technologies  
catalog #: 11547-015"  
BASE COUNT 169 a 145 c 120 g 178 t  
ORIGIN

Query Match 68.2%; Score 607.8; DB 10; Length 612;  
Best Local Similarity 99.7%; Pred. No. 3.4e-116;  
Matches 609; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 281 agaaaaaacacagaaaaaaatctcatatgcaaaatattccaccaggaacgaaga 340  
|||||  
Db 612 AGAAAAACCAACAGAAAAAATCTCATCATGCGCAAAATATTCACAGGAACGAAGA 553  
Qy 341 gatgagagcctatgcagaatggaggaagacccctttggagaggtgaaggcca 400  
|||||  
Db 552 GATGAGCAGCCTATGCGAATGGAGAGGAACCGCCCTTTGGGAGGAGGTGAAGGCCA 493  
Qy 401 ccagcctcaggaatcagcgggagcgcctcgcgcagcttgccttaatttcgatggc 460  
|||||  
Db 492 CCAGCCTCGAGGAATCGACGGGACAGCTCGCCGACTTGCCTTAATTTTCGATGGC 433  
Qy 461 catacccaatggcagatcaatgatggatggatggatggatggatggatggatggat 520  
|||||  
Db 432 CATACCAATAGGCAGATCAATGATGGATGGATGGATGGATGGATGGATGGATGAT 373  
Qy 521 catgagagatgagagaaatcagaagaaacttaggagctgcaggtgaggaattgtct 580  
|||||  
Db 372 CATGAGGAGATGAGAGAAATCAGAAAGAACTTAGGAGCTGCAGATTGAGGAATTGCT 313  
Qy 581 gcgtatcctatggggagctctcaatcacatgacatcatgatgaatttgccttat 640  
|||||  
Db 312 GCGTATCTTATGGGGAGCTCTTAATCACCATGACCATCAATGATGAATTTGCGCTTAT 253  
Qy 641 gcttgactcctgcctattatcatgagataactatgctgattccgcgtgtttcttttc 700  
|||||  
Db 252 GCCTTGACTCTGCCATTTATCATGAGATTAATACTGTGATTCGCCGTGTTTCTTTTC 193  
Qy 701 ctgtatttcttaataatgcctttactgatccgtttgtgtgaacccctatgtatttcca 760  
|||||  
Db 192 CTGTGATTTCTTAATATGCTTTTACTGATCGGTTGCTGTGAACCCCTATGTTATTTCCA 133  
Qy 761 ttgtcaagtgggtctgtgtgcagctctctattgagattgccttgcactcagtg 820  
|||||  
Db 132 TGTGTCAAGTGGCTCTGTGTGGCCAGCTTCTATTGTGAAGATTGCTTTGCACTCAGTGT 73  
Qy 821 aagtttctgcagcagtgatttccaccatttgcattgagaaaaatttaagccaataaagca 880  
|||||  
Db 72 AAGTTTCTGTACGAGTAGTTTCACCCATTTGCATGGAAAAATTTAAAGCTAATAAGCA 13  
Qy 881 atttaaaagc 891  
|||||  
Db 12 ATTTAAAAAGC 2

RESULT 2  
AL193112/c  
LOCUS AL193112 615 bp mRNA EST 29-OCT-1998  
DEFINITION qe69g09.x1 Soares\_fetal\_lung\_NbHL19W Homo sapiens cDNA clone  
IMAGE:1744288 3' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0  
KD PROTEIN HGR74 (HUMAN); mRNA sequence.  
ACCESSION AL193112  
VERSION AL193112.1 GI:3744321  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 615)  
AUTHORS NCICGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 845 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 445.  
Location/Qualifiers  
1. 615

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1744288"  
/clone\_lib="Soares\_fetal\_lung\_NbHL19W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAAGTGGAGGCGCGCAATTTTCTTTT-3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NBHL19W."  
BASE COUNT 169 a 145 c 121 g 180 t  
ORIGIN

Query Match 67.7%; Score 603; DB 10; Length 615;  
Best Local Similarity 99.2%; Pred. No. 3.3e-115;  
Matches 606; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 281 agaaaaaacacagaaaaaaatctcatatgcaaaatattccaccaggaacgaaga 340  
|||||  
Db 615 AGAAAAACCAACAGAAAAAATCTCATGATGGCAATATTCACAGGAACGAAGA 556  
Qy 341 gatgagcagcctatgcagaatggaggaagacccctttggagaggtgaaggcca 400  
|||||  
Db 555 GATGAGCAGCCTATGCGAATGGAGAGGAACCGCCCTTTGGGAGGAGGTGAAGGCCA 496  
Qy 401 ccagcctcaggaatcagcgggagcgcctcgcgcagcttgccttaatttcgatggc 460  
|||||  
Db 495 CCAGCCTCGAGGAATCGACGGGACAGCTTCGCCACTTGCCTTAATTTTCGATGGC 436  
Qy 461 catacccaatggcagatcaatgatggatggatggatggatggatggatggatggat 520  
|||||  
Db 435 CATACCAATAGGCAGATCAATGATGGATGGATGGATGGATGGATGGATGGATGAT 376  
Qy 521 catgagagatgagagaaatcagaagaaacttagggagctgcaggtgaggaattgtct 580  
|||||  
Db 375 CATGAGGAGATGAGAGAAATCAGAAAGAACTTAGGAGCTGCAGTTGAGGAATTGCTCT 316  
Qy 581 gcgtatcctatggggagcctctcaatcacatgacatcatgatgaatttgccttat 640  
|||||  
Db 315 GCGTATCTTATGGGGAGCTCTTAATCACCATGACCATGATGAATTTGCGCTTAT 256  
Qy 641 gcttgactcctgcctattatcatgagataactatgctgattccgcgtgtttcttttc 700  
|||||  
Db 255 GCCTTGACTCTGCCATTTATCATGAGATTAATACTGTGATTCGCCGCTGTTTCTTTTC 196  
Qy 701 ctgtcatttcttaataatgcctttactgatccgtttgtgtgaaacccctatgtatttcca 760  
|||||  
Db 195 CTGTCAATTTCTTAATATGCGCTTTACTGATCGTGTGCTGTGAACCCATGATTTATTTCCA 136  
Qy 761 ttgtcaagtgggtctgtgtgcccagctctctattgagattgccttgcactcagtg 820  
|||||  
Db 135 TGTGTCAAGTGGGTCTGTGTGGCCAGCTTCTATTGTGAAGATTGCTTTGCACTCAGTGT 76

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2002, 16:09:50 ; Search time 1609.39 Seconds  
(without alignments)  
5949.141 Million cell updates/sec

Title: US-09-327-750D-29  
Perfect score: 891  
Sequence: 1 accccatccccactctat.....aataaagcaatttaaaagc 891

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estfun:  
2: em\_esthum:  
3: em\_estin:  
4: em\_estom:  
5: em\_estpl:  
6: em\_estba:  
7: em\_estro:  
8: em\_estov:  
9: em\_htc:  
10: gb\_estl:  
11: gb\_est2:  
12: gb\_htc:  
13: gb\_gss:  
14: em\_gss\_fun:  
15: em\_gss\_hum:  
16: em\_gss\_inv:  
17: em\_gss\_pln:  
18: em\_gss\_pro:  
19: em\_gss\_rod:  
20: em\_gss\_vrt:  
21: em\_gss\_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	607.8	68.2	612	10 AW512400	AW512400 xx74b09.x
C 2	603	67.7	615	10 AI193112	AI193112 q669g09.x
C 3	597.2	67.0	626	10 AI928835	AI928835 au65b10.x
C 4	596.4	66.9	620	10 AA576958	AA576958 nm82d04.s
C 5	577.4	64.8	586	11 BF058872	BF058872 7k35g02.x
C 6	577	64.8	606	10 AI937016	AI937016 w672b04.x
C 7	574	64.4	662	10 AI929106	AI929106 au65b10.y
C 8	565.2	63.4	603	10 AA514324	AA514324 nf56b04.s
C 9	564.4	63.3	568	10 AW131584	AW131584 xf31g01.x
C 10	564.4	63.3	574	10 AI075989	AI075989 ov47b05.x
C 11	564.4	63.3	610	10 AV709328	AV709328 AV709328
C 12	562.8	63.2	574	10 BE222704	BE222704 hu51f06.x

13	562.8	63.2	745	10	AU132971	AU132971	AU132971	AU132971
C 14	562.4	63.1	566	10	AI598179	AI598179	AI598179	AI598179
C 15	560.8	62.9	566	11	BF732486	BF732486	BF732486	BF732486
C 16	560.8	62.9	720	10	AV701506	AV701506	AV701506	AV701506
C 17	560.4	62.9	566	10	AI859419	AI859419	AI859419	AI859419
C 18	559.8	62.8	584	10	AI188510	AI188510	AI188510	AI188510
C 19	559.6	62.8	837	10	BE728687	BE728687	BE728687	BE728687
C 20	557.8	62.6	578	10	AW156951	AW156951	AW156951	AW156951
C 21	557.8	62.6	599	10	AI033984	AI033984	AI033984	AI033984
C 22	555.2	62.3	606	10	AW088350	AW088350	AW088350	AW088350
C 23	555.2	62.3	779	11	BF764807	BF764807	BF764807	BF764807
C 24	553.6	62.1	676	11	BE896163	BE896163	BE896163	BE896163
C 25	553.4	62.1	716	11	BF716509	BF716509	BF716509	BF716509
C 26	552.4	62.0	573	11	BF940996	BF940996	BF940996	BF940996
C 27	551.8	61.9	772	11	BF339921	BF339921	BF339921	BF339921
C 28	550.8	61.8	858	11	BE793692	BE793692	BE793692	BE793692
C 29	550.4	61.8	690	11	BF825224	BF825224	BF825224	BF825224
C 30	549.8	61.7	589	10	AI188575	AI188575	AI188575	AI188575
C 31	549.6	61.7	556	10	AA573181	AA573181	AA573181	AA573181
C 32	549.6	61.7	728	11	BF972848	BF972848	BF972848	BF972848
C 33	549.4	61.7	566	10	AI033728	AI033728	AI033728	AI033728
C 34	548.8	61.6	736	10	BE408639	BE408639	BE408639	BE408639
C 35	548.8	61.6	879	11	BF338059	BF338059	BF338059	BF338059
C 36	548.6	61.6	701	11	BF403499	BF403499	BF403499	BF403499
C 37	548	61.5	674	11	BF745809	BF745809	BF745809	BF745809
C 38	547.6	61.5	775	11	BF706329	BF706329	BF706329	BF706329
C 39	546	61.3	707	10	BE732777	BE732777	BE732777	BE732777
C 40	545.4	61.2	704	11	BF488926	BF488926	BF488926	BF488926
C 41	543	60.9	570	10	AI887790	AI887790	AI887790	AI887790
C 42	540.4	60.7	730	11	BE900237	BE900237	BE900237	BE900237
C 43	538.2	60.4	621	10	AI570521	AI570521	AI570521	AI570521
C 44	537.4	60.3	574	10	AA843808	AA843808	AA843808	AA843808
C 45	536.6	60.2	682	10	BE727430	BE727430	BE727430	BE727430

## ALIGNMENTS

RESULT 1  
AW512400/c

LOCUS

DEFINITION  
xx74b09.x1 NCI\_CGAP\_Lym12 Homo sapiens cdna clone IMAGE:2849369 3', similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION  
AW512400

VERSION  
AW512400.1

KEYWORDS  
EST.

SOURCE  
human.

ORGANISM  
Homo sapiens

REFERENCE  
1 (bases 1 to 612)

AUTHORS  
NCI-CGAP

TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL  
Unpublished (1997)

COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Life Technologies catalog #: 11547-015

DNA Sequencing by: Washington University Genome Sequencing Center

Cloud distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LML at:

image.llnl.gov/image/html/iresources.shtml

Seq primer: -40up from Gibco

High quality sequence stop: 422.

Location/Qualifiers

1..612

Source

Organization="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2849369"

/clone\_lib="NCI\_CGAP\_Lym12"

/tissue\_type="lymphoma, follicular mixed small and large cell"



```
alignment_block:
US-09-327-750D-30 x AAF21748
Align seg 1/1 to: AAF21748 from: 1 to: 917
18 GlnLysGluGluLysGluLysGluLysProGlnAspThrIleArgArgG1 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 CAGGAAACCAAGAGATGGAGCGCTATCGAGAATCGAGAGGAGACCG 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 uProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 CCCTTTG.....GGAGGAGGTGAAGCCACCGCT...G 387
51 lYGLyArgArgArgPheArgValArgGlnProIleAlaHisTyrArgTrp 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 CAGAAATCAGCGGGACGCTGCGGACTTGCCTTAATTTTCGATGG 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 AspLeuMetGlnArg...ValGlyGluProGlnGlyArgMetArgGluG1 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 GCCATACCCCAATAGGCAGATCAATGATGGGATGGGTGA...GATGAGA 484
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 uAsnValGlnArgPheGlyClyAspValArgGlnLeuMetGluLysLeuA 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
485 TGATATGGAATATTCATGAGGAGATGAGAGAAATCAGAAGAAACTTA 534
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 rGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp...Pro 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
535 GGGAGCTGCAGTTCAGGAATGTCTGCGTATCCTTATGGGGGAGCTCT 584
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 ProHisAspHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
585 AATCACCATCACCATCATGATGAATTTGCTTATGCCT 623
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAC01005
seq_documentation_block:
ID AAC01005 standard; cDNA; 532 BP.
XX AAC01005;
AC AAC01005;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1003.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG00999.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 1003; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
```

```
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 532 BP; 151 A; 121 C; 163 G; 91 T; 6 other;

alignment_scores:
Quality: 119.50 Length: 115
Ratio: 1.707 Gaps: 5
Percent Similarity: 60.870 Percent Identity: 31.304

alignment_block:
US-09-327-750D-30 x AAC01005
Align seg 1/1 to: AAC01005 from: 1 to: 532
9 LysAsnLeuAsnMetGluAsnAspHisGlnLysLysGluLysGluG1 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 AAAAATCTCATTCATGGCAATATTCACGAGAAACGAGAGATGGAGCA 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
25 uLysProGlnAspThrIleArgArgGluProAlaValAlaLeuIleSerG 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
257 GCCTATGCAGATGGAGAGGAGAGACCCCTTTG.....G 291
42 luAlaGlyLysAsnCysAlaProArgGlyGlyArgArgPheArgVal 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 GAGGAGGTGAAGGCCACCGCCT...GCAGGAATCGACGGGAMAGGCT 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 ArgGlnProIleAlaHisTyrArgTrpAspLeuMetGlnArg...ValG1 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 CGCCGAYTTGCCCTAATTTTCGATGGCCATACCCAATAGGCAGATCAA 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
74 yGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheGlyGlyA 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 TGATGGGATGGGT...GRAGATGGAGATGATATGGAAATATTCATGGAGG 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 spValArgGlnLeuMetGluLysLeuArgGluArgGlnLeuSerHisSer 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 AGATGAGAGAAATCAGAAGAAACTTAGGAGCTGCAKTTTGAGGAATGT 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 LeuArg...AlaValSerThrAspProHisHisAspHisHis 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
486 CTGCGTATCCTTATGGGGGAKCTCTCTAATCACCATGACCATCAT 530
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```



```

50 ArgGlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrAr 66
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
329 GGAGGAATATCAGCGGGCGGAGTAGCGACTGTCCCTAATTTTCG 378
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 gTrpAspLeuMetGlnArg...ValGlyGluProGlnGlyArgMetArg 82
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 ATGGGCCATACCTAATAGCATATTGAGCACAATGAAGCGAGA..... 421

82 luGlusnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLys 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 ..GATGATGTAGAAAGGTTTGTAGGCAGCATGATGGAATCAAGAGAAAG 469

99 LeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
470 ACTAGGGAACAGCAGATAGGCACTATATGCGCTTCCAACTCCTGACCC 519

115 oProHisHisAspHisAspGluPheCysLeuMetPro 128
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
520 T.....GACAACCATATGACTTTTGCCTCATACCT 550

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT: AAC85548

seq\_documentation\_block:

ID AAC85548 standard; cDNA: 1364 BP.

AC AAC85548;

DT 04-JUN-2001 (first entry)

DE cDNA encoding CDIFF-4, Incyte ID No. 1990956CB1.

KW Human; cell differentiation; CDIFF; agonist; antagonist; epilepsy;  
 KW cell proliferation; Alzheimer's disease; schizophrenia disorder;  
 KW arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.

OS Homo sapiens.

FH Key Location/Qualifiers  
 FT CDS 250..612  
 FT /\*tag= a  
 FT /product= "CDIFF-4"

PN WO200119860-A2.

PD 22-MAR-2001.

PF 14-SEP-2000; 2000WO-US25435.

PR 15-SEP-1999; 99US-0154140.

PR 06-DEC-1999; 99US-0169155.

XX (INCY-) INCYTE GENOMICS INC.

PI Tang YT, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;  
 PI Baughn MR, Lu DAM, Bandman O, Shin LL, Patterson C;

DR WPI; 2001-211447/21.

DR P-PSDB; AAB47126.

XX Isolated polypeptides and polynucleotides involved in cell  
 PT differentiation are used for treatment, prevention and diagnosis of  
 PT cell proliferative, developmental and neurological disorders e.g.  
 PT cancer and Alzheimer's disease -

XX Claim 5; Page 121; 137pp; English.

XX The sequences given in AAC85545-72 encode human polypeptides involved  
 CC in cell differentiation (CDIFF). CDIFF polypeptides and agonists of  
 CC these are used to treat a disease or condition associated with  
 CC decreased expression of functional CDIFF. An antagonist of CDIFF is  
 CC used to treat a disease or condition associated with over expression  
 CC of functional CDIFF. CDIFF polypeptides may be used for the treatment,

CC prevention and diagnosis of cell proliferative, developmental and  
 CC neurological disorders, such as Alzheimer's disease, schizophrenic  
 CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus  
 CC and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus  
 CC REX-3. This sequence maps to chromosome 1 within the interval from  
 CC 152.2 to 157.4 centimorgans, to chromosome 3 within the interval from  
 CC 157.4 to 158.0 centimorgans, and to the X chromosome within the interval  
 CC from 104.9 to 150.3 centimorgans.

SQ Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;

alignment\_scores:

Quality: 170.00 Length: 130  
 Ratio: 1.954 Gaps: 6  
 Percent Similarity: 66.923 Percent Identity: 34.615

alignment\_block:

US-09-327-750D-30 x AAC85548 ..

Align seg 1/1 to: AAC85548 from: 1 to: 1364

```

1 MetGluSerLysAspGln...GlyValLysAsnLeuAsnMetGluAsnAs 16
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 ATGGAGTCCAAAGAGAGAACTAGCGGCAACAATCTCAACGGGGAAAATGC 299

16 phiGlnLysLysGluGluLysGluGluLysProGlnAspThrIleArgA 33
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 CCAACAAGAAACGAAGGAGGAGGAGCGGCCCC.....ACGCAGAAATG 343

33 rgGluProAlaValAlaLeuSerGluAlaGlyLysAsnCysAlaPro 49
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
344 AGAGAATATCCGCCCATTTG.....GGAGGGGTGAAGGCCAGAACCT 387

50 ArgGlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrAr 66
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
388 GGAGGAATATCAGCGGGCGGCGAGTTAGCGCACTTGTCCCTAATTTTCG 437

66 gTrpAspLeuMetGlnArg...ValGlyGluProGlnGlyArgMetArg 82
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
438 ATGGGCCATACCTAATAGCATATTGAGCACAATGAAGCGAGA..... 480

82 luGluAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLys 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
481 ..GATGATGTAGAAAGGTTTGTAGGCAGCATGATGGAATCAAGAGAAAG 528

99 LeuArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPr 115
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
529 ACTAGGGAACAGCAGATAGGCACTATATGCGCTTCCAACTCCTGACCC 578

115 oProHisHisAspHisAspGluPheCysLeuMetPro 128
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
579 T.....GACAACCATATGACTTTTGCCTCATACCT 609

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT: AAC23528

seq\_documentation\_block:

ID AAF23528 standard; DNA; 700 BP.

XX XX

AC AAF23528;

XX XX

DT 22-MAR-2001 (first entry)

XX XX

DE Mouse NADE DNA.

XX XX

KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;

KW neurogenetic disease; NF-kappaB; ds.

XX Mus sp.

OS Mus sp.

XX WO200075278-A2.

XX 14-DEC-2000.

PD 14-DEC-2000.

**THIS PAGE BLANK (USPTO)**

Quality: 81.00 Length: 121  
Ratio: 1.397 Gaps: 4  
Percent Similarity: 47.934 Percent Identity: 23.967

## alignment\_block:

US-09-327-750D-30 x US-08-459-448A-6 ..

Align seg 1/1 to: US-08-459-448A-6 from: 1 to: 3624

```
14  GluAsnAspHisGlnLysLysGluGluLysGluGluLysProGlnAspTh 30
||||:||||| :||||:||||| :||||:||||| :||||:|||||
1026 GAACACACGACGACGCGGCCACCAACACACGACATCAACCCGCGTACCC 1075
30  rIleArgArgGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnC 47
||||:||||| :||| :||||:||||| :||||:|||||
1076 TGGCGCTTCGCCGCGCGGACGTGTACCGCACCGAGAGCTACGCCGCGGTG 1125
47  ysAla.....ProArgGlyArgArgArg 55
:||||| :||||| :||||:|||||
1126 CTGCTGTGGGGCATCTACCTGGAGCCCATCCAGCGGCGTGCCCAACCGTGG 1175
56  PheArgValArgGlnProIleAlaHisTyrArgTrpAspLeuMetGlnAr 72
||||:||||| :||| :||| :|||
1176 CTTCAACTTCACCAACCCCGACGACATCAGCGA.....CCGCG 1213
72  gValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnArgPheG 89
| :||| :||| :||| :|||
1214 GCACCCCAACTACACCCAGCCCTACGAGAGCCC..... 1247
89  lyGlyAspValArgGlnLeuMetGluLysLeuArgGluArg..... 102
||| :||| :||| :|||
1248 .....CGGCCTGCAGCTGAAGACAGCGAGACCGGCTGCCCCC 1286
103 .....GlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
||| :||| :||| :|||
1287 CGAGACCGCGCGCGCCCACTACGAGAGCTACAGCCACCGCGCTGAGCC 1336
117 sHisAspHisHis 121
||| :||| :||| :|||
1337 ACATCGGCATCAT 1349
```



120 sHisAspGluPheCysLeuMetPro 128  
 |||||  
 596 TCATGATGAGTTTGGCNWATGCC 620

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF23529

seq\_documentation\_block:

ID AAF23529 standard; DNA; 891 BP.

XX

AC AAF23529;

XX

DT 22-MAR-2001 (first entry)

XX

DE Human NADE DNA.

XX

KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;  
 neurogenetic disease; NF-kappaB; ds.

XX

OS Homo sapiens.

XX

PN WO200075278-A2.

XX

PD 14-DEC-2000.

XX

PF 07-JUN-2000; 2000WO-US15621.

XX

PR 07-JUN-1999; 99US-0327750.

XX

PA (UYCO ) UNIV COLUMBIA NEW YORK.

XX

PI Sato T;

XX

DR WPI; 2001-061707/07.

XX

PT New p75-neurotrophin receptor-associated cell death executor (NADE) and  
 the gene encoding NADE, useful for modulating the activity of p75NTR  
 and for detecting neurodegenerative diseases .

XX

PS Disclosure; Fig 1; 134pp; English.

XX

CC The present invention relates to a purified polypeptide capable of  
 binding neurotrophin receptor (p75-NTR). The invention is useful for  
 binding and modulating the activity of p75NTR. The peptide mediates  
 CC NGF-induced apoptosis, which plays an important role in neurogenetic  
 CC diseases. The peptide of the invention and p75NTR are useful for  
 CC inhibiting NF-kappaB activation in a cell or a subject, for inducing  
 CC caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase  
 CC and fragment nuclear DNA in a cell by co-expression of (I) and p75-NTR.

XX

SQ Sequence 891 BP; 251 A; 182 C; 224 G; 234 T; 0 other;

alignment\_scores:

Quality: 171.50 Length: 129

Ratio: 2.091 Gaps: 5

Percent Similarity: 63.566 Percent Identity: 34.884

alignment\_block:

US-09-327-750D-30 x AAF23529 ..

Align seg 1/1 to: AAF23529 from: 1 to: 891

2 GluSerLysAspGlnGluValLysAsnLeuAsnMetGluAsnAspHisG1 18

|||||

279 AAGAGAAAACACAGAAAAAAATCTCATGCGCAATATTCACCA 328

|||||

18 nLysLysGluGluLysGluGluLysProGlnAspThrIleArgGluP 35

|||||

329 GGAACACGAGATGGAGCGCTATGCGAATGGAGAGACCGCC 378

|||||

35 roAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArgGly 51

|||||

|||||

379 CTTTG.....GGAGGAGGTGAAGGCCACCGCCT...GCA 410

52 GlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTTPas 68

|||||

411 GGAATCGACGGGACAGGCTCGCCGACTTGCCTTAATTTTCGATGGC 460

|||||

68 pLeuMetGlnArg...ValGlyGluProGlnGlyArgMetArgGluGluA 84

|||||

461 CATACCAATAGGCAGATCAATGATGGATGGGTGA...GATGGAGATG 507

|||||

84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100

|||||

508 ATATGGAATATTCATGGAGGAGATGAGAAATCAGAAGAAACTTAGG 557

|||||

101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAsp...ProPr 116

|||||

558 GAGCTGCAGTTGAGGAATGTCTGCTATCTTATGGGAGGCTCTCTAA 607

|||||

116 oHisHisAspHisAspGluPheCysLeuMetPro 128

|||||

608 TCACCATGACCATCATGATGAATTTTCCTTATGCCT 644

|||||

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH03517

seq\_documentation\_block:

ID AAH03517 standard; cDNA; 865 BP.

XX

AC AAH03517;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA clone (5'-primer) SEQ ID NO:352.

XX

KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-0116126.

XX

PR 29-JUL-1999; 99JP-0248036.

XX

PR 27-AUG-1999; 99JP-0300253.

XX

PR 11-JAN-2000; 2000JP-0118776.

XX

PR 02-MAY-2000; 2000JP-0183767.

XX

PR 09-JUN-2000; 2000JP-0241899.

XX

PA (HELI-) HE'LY RES INST.

XX

PI Ota T, Isoyui T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX

PI Ishii S, Sogiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-3\*8749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602

XX

PT full-length cDNAs defined in the specification, and for the detection

XX

PT and/or diagnosis of the abnormality of the proteins encoded by the

XX

PT full-length cDNAs .

XX

PS Claim 1; SEQ ID 352; 2537pp + CD ROM; English.

XX

CC The present invention describes primer sets for synthesizing 5602

XX

CC full-length cDNAs defined in the specification. Where a primer set

XX

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

XX

CC to the complementary strand of a polynucleotide which comprises one of

XX

CC the 5602 nucleotide sequences defined in the specification, where the

XX

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX

CC of an oligonucleotide comprising a sequence complementary to the

XX

CC complementary strand of a polynucleotide which comprises a 5'-end

XX

CC sequence and an oligonucleotide comprising a sequence complementary to a

XX

CC polynucleotide which comprises a 3'-end sequence, where the

XX





CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities, chemotactic/chemokinetic activity, haemostatic  
CC Activin/inhibin activity, cancer diagnosis and therapy, drug screening, and  
CC and thrombolytic activity, arthritis and inflammation, leukaemias and  
CC assays for receptor activity, C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 858 BP; 236 A; 175 C; 241 G; 206 T; 0 other;

alignment\_scores:  
Quality: 437.00 Length: 130  
Ratio: 3.767 Gaps: 3  
Percent Similarity: 89.231 Percent Identity: 66.154

alignment\_block:  
US-09-327-750D-30 x AAC160367 ..  
Align seg 1/1 to: AAC160367 from: 1 to: 858

1 MetGluSerLysAspGln...GlyValLysAsnLeuAsnMetGluAsnAs 16  
|||||  
215 ATGGAGTCCAAAGAGGACGCGTTAAACAATCTCATCGTGGAAAATGT 264  
16 pHisGlnLysLysGluLysGluGluLysProGlnAspThrIleArgA 33  
|||||  
265 CAACCGAGAAATGATGAAAGATGAAAGGAGCAAGTTGCTTAATAAAG 314  
33 rGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaPro 49  
|||||  
315 GGGAGCCC...TTGGCCCTACCTTTGATGTTAGTGAATACTGTGCGCT 361  
50 ArgGlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrAr 66  
|||||  
362 AGAGGAACCGTAGCGGTTCCGCGTTAGGACGCCCATCTCGCAGTATAG 411  
66 gTrpAspLeuMetClnArgValGlyGluProGlnGlyArgMetArgGluC 83  
|||||  
412 ATGGGACATTAATCATAGGCTTGGAGGCCACAGGCAAGGATGAGAGG 461  
83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99  
|||||  
462 AGAATATGGAAGGATGCGGAGGAGGTGAGACAGCTGATGGAAGCTG 511  
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116  
|||||  
512 AGGGAAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCACTGATCCCC 561  
116 oHisHisAspHisHisAspGluPheCys.LeuMetPro 128  
|||||  
562 TCACCATGACCATCAGATGAGTTTGGCCCTTATGCC 599

seq\_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT: AAC10889

seq\_documentation\_block:  
ID AAC10889 standard; cDNA: 692 BP.

XX  
AC AAC10889;

XX  
DT 06-OCT-2000 (first entry)

XX  
DE Human secreted protein 5' EST, SEQ ID NO: 14964.

XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

XX  
OS Homo sapiens.

XX  
PN EP1033401-A2.

XX

PD 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-0200610.  
XX  
XX 26-FEB-1999; 99US-0122487.  
XX  
XX (GEST ) GENSET.  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
XX WPI: 2000-500381/45.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 14964; 71pp + CD-ROM; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively  
XX identified within the present sequence. The 5' ESTs were prepared from  
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
XX sequences usually correspond mainly to the 3' untranslated region (UTR)  
XX of the mRNA because they are often obtained from oligo-dT primed cDNA  
XX libraries. Such ESTs are not well suited for isolating cDNA sequences  
XX derived from the 5' ends of mRNAs and even in those cases where longer  
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.  
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX They are used to obtain upstream regulatory sequences and to design  
XX expression and secretion vectors.  
XX  
XX Sequence 692 BP; 184 A; 151 C; 199 G; 146 T; 12 other;

alignment\_scores:  
Quality: 429.50 Length: 125  
Ratio: 3.940 Gaps: 1  
Percent Similarity: 87.200 Percent Identity: 64.000

alignment\_block:  
US-09-327-750D-30 x AAC10889 ..  
Align seg 1/1 to: AAC10889 from: 1 to: 692

4 LysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHisGlnLysLy 20  
|||||  
249 AGAGAACGAGCGTTAAACAATCTCATCGTGGAAAATGTCAACCCAGGAAAA 298  
20 sGluGluLysGluGluLysProGlnAspThrIleArgArgGluProAlav 37  
|||||  
299 TGATGAAAAGATGAAAAGGAGCAAGTGWCTAATAAAGGGGAGGCC...T 345  
37 alAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArgGlyGlyArg 53  
|||||  
346 TGGCCCTACCTTTGATGTYTRGTGATACTACTGTGCTAGAGGAATCGT 395  
54 ArgArgPheArgValArgGlnProIleAlaHisTyrArgTrpAspLeuMe 70  
|||||  
396 AGGCGGTTCCGCTTAGGAGGCCCATCTGCAGTATAGATGGATATGAT 445  
70 tGlnArgValGlyGluProGlnGlyArgMetArgGluGluAsnValGlnA 87  
|||||  
446 GCATAGGCTTGGAGAACCAAGGAGGATGANAGAGAGATATATGGAAA 495  
87 rGpPheGlyGlyAspValArgGlnLeuMetGluLysLeuArgGluArgGln 103  
|||||  
496 GGATTGGGGAGGAGGTGACACAGCTGATGGAAGCTGAGGGAAGACGAG 545  
104 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHis 120  
|||||  
546 TTGAGTCATAGTCTGGGGCAGTCAGCACTGACCCCTCACCATTGACCA 595





```
XX SQ Sequence 1229 BP; 351 A; 243 C; 296 G; 339 T; 0 other;
alignment_scores:
  Quality: 264.50      Length: 123
  Ratio: 3.112        Gaps: 4
  Percent Similarity: 69.106 Percent Identity: 47.967
alignment_block:
  US-09-327-750D-12 x AAH13750
  Align seg 1/1 to: AAH13750 from: 1 to: 1229
    3 AsnValHisGlnGluAsnGluMetGluGln...ProLeuGlnAsnG1 18
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    236 AATGCCCAACAAGAAACAGAGGAGGAGGAGGCCGCCACGCGAATGA 285
    18 yGluGluAspArgProValGlyGlyGlyGlyGlyGlyHisGlnProAlaGlyA 35
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    286 AGAAGAATCCCGCATTTGGGAGGGGTGAAGGCCAGAGCCTGGAGGAA 335
    35 snAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArgArgGly 51
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    336 AT.....ATCAGCGCGGGG 349
    52 GlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnArgG1 68
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    350 CGAGTAGGCGACTTCCCTAATTTTCGATGGCGCATACCTAATAGGCA 399
    68 nMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyMetPheMetG 85
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    400 TATTGACCAACAAGCG.....AGAGATGATGTAGAAAGTTTCTAG 443
    85 luGluMetArgGluIleArgArgGlyLeuArgGlyLeuGluLeuArgAsn 101
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    444 GGCAGATGATGGAATCAAGAGAAGACTAGGGAACAGCAGATGAGGCAC 493
    102 CysLeuArgIleLeuMetGlyGlyLeuLeuSerAsnHisHisHisHis 118
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    494 TATATGCGCTTCCAACCTGTAACCTGACCAACCATATTATGAC..... 535
    118 pGluPheCysLeuMetPro 124
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    536 .....TTTTCCTCATACCT 550
seq_name: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT:AAc85548
seq_documentation_block:
  ID AAC85548 standard; cDNA; 1364 BP.
XX AC AAC85548;
XX DT 04-JUN-2001 (first entry)
XX DE cDNA encoding CDIFF-4, Incyte ID No. 1990956CB1.
XX KW Human; cell differentiation; CDIFF; agonist; antagonist; epilepsy;
XX cell proliferation; Alzheimer's disease; schizophrenia disorder;
XX arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 250..612
XX FT /*tag= a
XX FT /product= "CDIFF-4"
XX PN WO200119860-A2.
XX PD 22-MAR-2001.
XX PF 14-SEP-2000; 2000WO-US25435.
```

```
XX 15-SEP-1999; 99US-0154140.
PR 06-DEC-1999; 99US-0169155.
XX (INCY-) INCYTE GENOMICS INC.
XX Tang YT, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;
PI Baughn MR, Lu DAM, Bandman O, Shih LL, Patterson C;
XX WPI; 2001-211447/21.
DR P-PSDB; AAB47126.
XX Isolated polypeptides and polynucleotides involved in cell
PT differentiation are used for treatment, prevention and diagnosis of
PT cell proliferative, developmental and neurological disorders e.g.
PT cancer and Alzheimer's disease -
XX Claim 5; Page 121; 137pp; English.
XX The sequences given in AAC85545-72 encode human polypeptides involved
CC in cell differentiation (CDIFF). CDIFF polypeptides and agonists of
CC these are used to treat a disease or condition associated with
CC decreased expression of functional CDIFF. An antagonist of CDIFF is
CC used to treat a disease or condition associated with over expression
CC of functional CDIFF. CDIFF polypeptides may be used for the treatment,
CC prevention and diagnosis of cell proliferative, developmental and
CC neurological disorders, such as Alzheimer's disease, schizophrenic
CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus
CC and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus
CC REX-3. This sequence maps to chromosome 1 within the interval from
CC 152.2 to 157.4 centiMorgans, to chromosome 3 within the interval from
CC 157.4 to 158.0 centiMorgans, and to the X chromosome within the interval
CC from 104.9 to 150.3 centiMorgans.
XX SQ Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;
alignment_scores:
  Quality: 264.50      Length: 123
  Ratio: 3.112        Gaps: 4
  Percent Similarity: 69.106 Percent Identity: 47.967
alignment_block:
  US-09-327-750D-12 x AAC85548
  Align seg 1/1 to: AAC85548 from: 1 to: 1364
    3 AsnValHisGlnGluAsnGluGluMetGluGln...ProLeuGlnAsnG1 18
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    295 AATGCCCAACAAGAAACAGAGGAGGAGGAGGCCGCCACGCGAATGA 344
    18 yGluGluAspArgProValGlyGlyGlyGlyGlyGlyHisGlnProAlaGlyA 35
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    345 AGAAGAATCCCGCATTTGGGAGGGGTGAAGGCCAGAGCCTGGAGGAA 394
    35 snAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHisArgArgGly 51
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    395 AT.....ATCAGCGCGGGG 408
    52 GlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnArgG1 68
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    409 CGAGTAGGCGACTTGCCTAATTTTCGATGGCGCATACCTAATAGGCA 458
    68 nMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyMetPheMetG 85
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    459 TATTGACCAACAAGCG.....AGAGATGATGTAGAAAGTTTCTAG 502
    85 luGluMetArgGluIleArgArgGlyLeuArgGlyLeuGluLeuArgAsn 101
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    503 GGCAGATGATGGAATCAAGAGAAGACTAGGGAACAGCAGATGAGGCAC 552
    102 CysLeuArgIleLeuMetGlyGlyLeuLeuSerAsnHisHisHisHis 118
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

XX (HELI-) HELIX RES INST.  
PA Ota T, Isogai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 1; SEQ ID 352; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
XX  
XX Sequence 865 BP; 239 A; 177 C; 224 G; 221 T; 4 other;

## alignment\_scores:

Quality: 264.50 Length: 123  
Ratio: 3.112 Gaps: 4  
Percent Similarity: 69.106 Percent Identity: 47.967

## alignment\_block:

US-09-327-750D-12 x AAH03517 ..

Align seg 1/1 to: AAH03517 from: 1 to: 865

3 AsnValHisGlnGluAsnGluMetGluGln...ProLeuGlnAsnG1 18  
236 AATGCCCAACAGAAACAGGAGGGGAGAGGCCGCCACGCGAATGA 285  
18 yGluGluAspArgProValGlyGlyGluGlyHisGlnProAlaGlyA 35  
286 AGAAGAAATCCCGCATTTGGAGGGGTGAGGCCAGAACCTTGAGGAA 335  
35 snAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArgArgGly 51  
336 AT.....ATCAGCGCGGGG 349  
52 GlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnArg1 68  
350 CGAGTTAGGCGATCTTCCTTAATTTTCGATGGCCATACCTAATAGGCA 399  
68 nMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheMetG 85  
400 TATTGAGCACATGAACGG.....ACAGATGATGTAGAAAGCTTTGTAG 443  
85 GluMetArgGluLeuLeuArgArgLysLeuArgGluLeuGlnLeuArgAsn 101  
.....

444 GGCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATGAGGCAC 493  
102 CysLeuArgIleLeuMetGlyLeuSerAsnHisHisAspHisHisAs 118  
494 TATATGCGCTTCCAAACTCTGAACTGACACCATTTATGAC..... 535  
118 pGluPheCysLeuMetPro 124  
536 ....TTTTCCTCATACCT 550  
seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH13750  
seq\_documentation\_block:  
ID AAH13750 standard; cDNA; 1229 BP.  
XX  
XX AC AAH13750;  
XX  
XX 26-JUN-2001 (first entry)  
XX  
XX Human cDNA sequence SEQ ID NO:10656.  
XX  
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
XX  
XX 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
XX  
XX 27-AUG-1999; 99JP-0300253.  
XX  
XX 11-JAN-2000; 2000JP-0118776.  
XX  
XX 02-MAY-2000; 2000JP-0183767.  
XX  
XX 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 10656; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.



```

|||||
269 TGCAGAGGAAGACCGCCCTTTGGGAGGAGGTGAAGCCACCACCGCTGCAG 318
|||||
34  lyAsnAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArg 50
|||||
319 GAAAT.....CGACGG 329
|||||
51  GlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAlaIleProAsnAr 67
|||||
330 GGAMAGCTCGCGGAYTTGCCCTTAATTTTCGATGGCCATACCCAATAG 379
|||||
67  gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
|||||
380 GCAGATCAATGATGGTGGTGRAGATGGATGATATGGAATAATCA 429
|||||
84  etGluGluMetArgGluIleArgArgLysLeuArgGluLeuLeuArg 100
|||||
430 TGGAGGAGATGAGAGAAATCAGAAGAAACTTAGGGAGCTGCAKTTGAGG 479
|||||
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisAspHisH 117
|||||
480 AATGTCTGGTATCTTATGGGGAKCTCTCTAATCACCATGACCATCA 529
|||||
117 s 117
|||||
530 T 530

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH01004

seq\_documentation\_block:

ID AAC01004 standard; cDNA; 485 BP.

AC AAC01004;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 1002.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

PF 26-FEB-1999; 99US-0122487.

PR (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR P-PSDB; AAG00998.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 1002; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.

XX Sequence 485 BP; 125 A; 113 C; 159 G; 84 T; 4 other;

alignment\_scores:

Quality: 489.50 Length: 112  
 Ratio: 5.099 Gaps: 1  
 Percent Similarity: 85.714 Percent Identity: 81.250

alignment\_block:

US-09-327-750D-12 x AAC01004 ..

Align seg 1/1 to: AAC01004 from: 1 to: 485

```

6  GlnGluAsnGluMetGluGlnProLeuGlnAsnGlyGluGluAspAr 22
|||||
187 CAGAAACAGAGAGATGGAGCAGCCCTATGCAGATGGAGGAGACCG 236
|||||
22  gProValGlyGlyGlyHisGlnProAlaGlyAsnAsnAsnAsnA 39
|||||
237 CCCTTTGGGAGGAGGTGAAGCCACCAGCTGCAGGAAT..... 276
|||||
39  snAsnHisAsnHisAsnHisAsnHisArgArgGlyGlnAlaArgArg 55
|||||
277 .....CGACGGGACAGCTCGCCGA 297
|||||
56  LeuAlaProAsnPheArgTTPAlaIleProAsnArgGlnMetAsnAspG 72
|||||
298 CTTGGCCCTAATTTTCGATGGCCATACCCAATAGGCAGATCAATGATGG 347
|||||
72  yLeuGlyGlyAspGlyAspMetGluMetPheMetGluGluMetArgG 89
|||||
348 GATGGGTGRAGATGGAGATGATATGGAAATATTCATGGAGAGATGAGAG 397
|||||
89  lulleArgArgLysLeuArgGluLeuGlnLeuArgAsnCysLeuArgIle 105
|||||
398 AAATCAGAGAGAAACTTAGGGAGCTGCAKTTGAGGAATTTGCTGCGTATC 447
|||||
106 LeuMetGlyGluLeuSerAsnHisAspHisHis 117
|||||
448 CTTATGGGGAKCTCTCTAATCACCATGACCATCAT 483

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH03517

seq\_documentation\_block:

ID AAH03517 standard; cDNA; 865 BP.

XX AAH03517;

XX 26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:352.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

```

XX WPI; 2000-611515/58.
DR P-PSDB; AAB58845.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases.
XX
XX Claim 1; Page 581-582; 1299pp; English.
XX
CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC antidiabetic; antiinflammatory; antiulcer; vulnerary; hepatotropic;
CC antibacterial; antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and agonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 917 BP; 228 A; 203 C; 270 G; 211 T; 5 other;

alignment_scores:
  Quality: 550.50      Length: 119
  Ratio: 5.193        Gaps: 1
  Percent Similarity: 89.076      Percent Identity: 84.874

alignment_block:
US-09-327-750D-12 x AAF21748 ..
Align seg 1/1 to: AAF21748 from: 1 to: 917
6 GlnGluAsnGluGluMetGluGlnProLeuGlnAsnGlyGluGluAspar 22
306 CAGGAAACCAAGAGATGGAGCAGCCTATGCAGATGGAGAGAACCG 355
22 gProValGlyGlyGlyGlyHisGlnProAlaGlyAsnAsnAsnAsn 39
356 CCCTTTGGGAGAGGTGAAGGCCACCAGCCTGCAGGAAT..... 395
39 snAsnHisAsnHisAsnHisAsnHisHisArgArgGlyGlnAlaArg 55
396 .....CGACGGGGACAGGCTGCCCGA 416
56 LeuAlaProAsnPheArgTrpAlaIleProAsnArgGlnMetAsnAsp 72
417 CTTGGCCCCCAATTTTCGATGGGCCATATCCCAATAGCAGATCAATG 466
72 yLeuGlyGlyAspGlyAspMetGluMetPheMetGluGluMetArg 89
467 GATGGGTGGAGATGGAGATGATATGGAATATTCATGGAGGAGATGAG 516
89 luIleArgArgLysLeuArgGluLeuGlnLeuArgAsnCysLeuArg 105
517 AAATCAGAAAGAACTTAGGAGCTCAGTTGAGGAATTTCTCTCGCTATC 566
106 LeuMetGlyGluLeuSerAsnHisHisAspHisHisAspGluPheCys 122
567 CTTATGGGGGAGCTCTCTAATCACCATGACCATCATGATGAATTTTGC 616
122 uMetPro 124

```

---

```

|||||||
617 TATGCCT 623

seq_name: /SID62/gcgdata/geneseq/geneseq/NA2000.DAT:AAC01005
seq_documentation_block:
ID AAC01005 standard; cDNA: 532 BP.
XX
AC AAC01005;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1003.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG00999.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 1003; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dr primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 532 BP; 151 A; 121 C; 163 G; 91 T; 6 other;

alignment_scores:
  Quality: 504.50      Length: 117
  Ratio: 5.096        Gaps: 1
  Percent Similarity: 84.615      Percent Identity: 79.487

alignment_block:
US-09-327-750D-12 x AAC01005 ..
Align seg 1/1 to: AAC01005 from: 1 to: 532
1 MetAlaAsnValHisGlnGluAsnGluGluMetGluGlnProLeuGlnAs 17
219 ATGGCAATATTATTCACCAAGAAACGAAGATGGAGCAGCCTATGCAGAA 268
17 nGlyGluGluAspArgProValGlyGlyGlyGlyHisGlnProAlaG 34

```



Tue Mar 12 09:01:16 2002

Date: Mar 11, 2002 3:42 PM

Command line parameters:

```
-MODEL=frameat_p2n_model -DEV=exlh
-Q/cqn2.1/USPO_spool/US09327750/runat_11032002_101154_20362/app_query.fasta_1.1472
-DB=N_Genesec_1101 -QMT=fastcap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINWATX=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODEL=LOCAL -OUTFMT=pfs
-NORM_ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09327750_@CGT_1_330 -NCPU=6 -ICPUX=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Query: US-09-327-750D-12

Database: N\_Genes

Database length: 428662619

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA23529+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA23529+		685.0	1330.8	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA23529+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA23529+		576.50	1115.8	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA21748+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA21748+		550.50	1064.5	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA201005+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA201005+		504.50	979.6	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA201004+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA201004+		489.50	951.1	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA1003517+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA1003517+		264.50	504.4	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA113750+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA113750+		264.50	501.0	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA835548+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA835548+		264.50	500.0	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA871747+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA871747+		193.00	379.2	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA451413+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA451413+		184.00	347.5	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA475810+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA475810+		171.50	323.3	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA158581+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA158581+		171.50	322.2	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA1959611+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA1959611+		171.50	321.8	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA825986+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA825986+		168.00	322.2	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA303880+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA303880+		162.50	307.1	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA3010889+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA3010889+		162.50	306.7	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA303879+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA303879+		162.50	306.6	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA160367+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA160367+		158.50	296.7	
/SIDS2/gcgdata/geneseg/geneseq/NA1999.DAT.AA190666+	/SIDS2/gcgdata/geneseg/geneseq/NA1999.DAT.AA190666+		101.00	174.6	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA3732130+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA3732130+		101.00	174.6	
/SIDS2/gcgdata/geneseg/geneseq/NA1998.DAT.AA563032+	/SIDS2/gcgdata/geneseg/geneseq/NA1998.DAT.AA563032+		101.00	172.5	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA4047157+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA4047157+		100.00	177.8	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA2001003+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA2001003+		99.50	188.5	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA4949183+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA4949183+		95.50	171.1	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA3089865+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA3089865+		95.50	170.3	
/SIDS2/gcgdata/geneseg/geneseq/NA1989.DAT.AA8991818+	/SIDS2/gcgdata/geneseg/geneseq/NA1989.DAT.AA8991818+		95.50	163.8	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA4949164+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA4949164+		95.50	161.1	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA4075288+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA4075288+		94.50	160.0	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA429736+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA429736+		92.50	156.4	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA078446+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA078446+		92.00	167.7	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA307325+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA307325+		90.00	169.7	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA472409+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA472409+		87.50	154.3	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA495484+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA495484+		86.50	147.5	
/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA887111+	/SIDS2/gcgdata/geneseg/geneseq/NA2000.DAT.AA887111+		86.50	147.5	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA006449+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA006449+		86.50	145.6	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA114082+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA114082+		85.00	160.1	
/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA135463+	/SIDS2/gcgdata/geneseg/geneseq/NA2001.DAT.AA135463+		85.00	160.1	
/SIDS2/gcgdata/geneseg/geneseq/NA1993.DAT.AA037948+	/SIDS2/gcgdata/geneseg/geneseq/NA1993.DAT.AA037948+		84.50	141.9	
/SIDS2/gcgdata/geneseg/geneseq/NA1998.DAT.AA0750330+	/SIDS2/gcgdata/geneseg/geneseq/NA1998.DAT.AA0750330+		84.50	141.9	







```
;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 663..3164
;
US-08-396-001-3

alignment_scores:
    Quality: 79.00      Length: 92
    Ratio: 1.362        Gaps: 5
    Percent Similarity: 63.043    Percent Identity: 26.087

alignment_block:
US-09-327-750D-12 x US-08-396-001-3 ..
Align seg 1/1 to: US-08-396-001-3 from: 1 to: 3455

17 AsnGlyGluGluAspArgProValGlyGlyGlyGlyGlyHisGlnProAl 33
2451 AACACGAAATAAGAACCC.....CATAACAAAAA 2482

33 aGlyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHisArgA 50
2483 TAGTCATAATCATATCATATCATATCATATCATATCATATCATAT 2532

50 rgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIlePro... 65
2533 ATAACATAATAATCAAAAGAGTCATACCCGTCATTTTCTTTACCAGCT 2582

66 .....AsnArgGlnMetAsnAspGlyLeuGlyGlyAspAs 79
2583 AATGCTTACCATAGAGAAGTACACGCTCTGTAAACCAATAAT..... 2624

79 pMetGluMetPheMetGluGlu..MetArgGluIleArgArgLysLeuArg 95
2625 .....TTCTCAACCAATATG.....CACAAATCA 2676

96 GluLeuGlnLeuArgAsnCysLeu 103
2653 AAATCACTCTCCGCAACAAATTA 2676

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-323-433A-3

seq_documentation_block:
; Sequence 3, Application US/09323433A
; Patent No. 6218512
;
GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James J.
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESCENCE IN
; FILE REFERENCE: 0050.1491-003
; CURRENT APPLICATION NUMBER: US/09/323.433A
; PRIOR FILING DATE: 1999-06-01
; PRIOR FILING DATE: 1995-02-28
; PRIOR APPLICATION NUMBER: PCT/US94/09351
; PRIOR FILING DATE: 1994-08-15
; PRIOR APPLICATION NUMBER: US 08/107,408
; PRIOR FILING DATE: 1993-08-16
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 3455
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (663)...(3164)
```

```
;
; OTHER INFORMATION: UTH4
;
US-09-323-433A-3

alignment_scores:
    Quality: 79.00      Length: 92
    Ratio: 1.362        Gaps: 5
    Percent Similarity: 63.043    Percent Identity: 26.087

alignment_block:
US-09-327-750D-12 x US-09-323-433A-3 ..
Align seg 1/1 to: US-09-323-433A-3 from: 1 to: 3455

17 AsnGlyGluGluAspArgProValGlyGlyGlyGlyGlyHisGlnProAl 33
2451 AACACGAAATAAGAACCC.....CATAACAAAAA 2482

33 aGlyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHisArgA 50
2483 TAGTCATAATCATATCATATCATATCATATCATATCATATCATAT 2532

50 rgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIlePro... 65
2533 ATAACATAATAATCAAAAGAGTCATACCCGTCATTTTCTTTACCAGCT 2582

66 .....AsnArgGlnMetAsnAspGlyLeuGlyGlyAspAs 79
2583 AATGCTTACCATAGAGAAGTACACGCTCTGTAAACCAATAAT..... 2624

79 pMetGluMetPheMetGluGlu..MetArgGluIleArgArgLysLeuArg 95
2625 .....TTCTCAACCAATATG.....CACAAATCA 2676

96 GluLeuGlnLeuArgAsnCysLeu 103
2653 AAATCACTCTCCGCAACAAATTA 2676

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-155-888-1

seq_documentation_block:
; Sequence 1, Application US/08155888
; Patent No. 6066623
;
GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Hedstrom, Richard C.
; APPLICANT: Sedegah, Martha
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE PROTECTIVE
; TITLE OF INVENTION: AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR
; TITLE OF INVENTION: DELIVERING POLYNUCLEOTIDE VACCINES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Naval Medical Res. & Dev. Cmd.
; STREET: Bldg. 1, T-12 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/155,888
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, A. David
; REGISTRATION NUMBER: 24,743
; REFERENCE/DOCKET NUMBER: N.C. 75.851
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 295-6759
```



```

Percent Similarity: 63.043 Percent Identity: 26.087
alignment_block:
US-09-327-750D-12 x US-08-861-464-3 ..
Align seg 1/1 to: US-08-861-464-3 from: 1 to: 3455

17 AsnGlyGluGluAspArgProValGlyGlyGluGlyHisGlnProAl 33
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2451 AACACGAAATAACGACCCC.....CATAACAAAA 2482

33 aGlyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArgA 50
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2483 TAGTCATATATCATATCATATATCATATCATATCAACCATGCTACAAATAA 2532

50 rgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIlePro... 65
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2533 ATACCAATAATAATCAAAAGAGTCATACCCGTCATTTTCTTTACCAGCT 2582

66 .....AsnArgGlnMetAsnAspGlyLeuGlyGlyAspGlyAspAs 79
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2583 AATGCTTACCATAAGAAGTAAACAGCTCTGTAAACCAATAAT..... 2624

79 pMetGluMetPheMetGluGlu.MetArgGluIleArgArgLysLeuArg 95
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2625 .....TTCTCAACCAATATG.....CACAAAGATCAGA 2652

96 GluLeuGlnLeuArgAsnCysLeu 103
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2653 AAATTCACCTCTCCGCAACAAATTA 2676

seq_name: /cgn2_6/ptodata/2/ins/5B_COMB.seq:US-08-396-001-3

seq_documentation_block:
; Sequence 3, Application US/083960001
; Patent No. 5919618
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence i
; TITLE OF INVENTION: Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,001
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6409A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

95 rgGluLeuGlnLeu.....ArgAsnCysLeu.....Arg 104
||||| :|||
280623 AGCAGCGGACGTTTCGGCGCATCAGTCCGCGTGGCGGAACCAACGCGC 280672
105 IleLeuMetGlyGluLeuSerAsnHisHis.....AspHis 116
||| :|||
280673 TTGCCACAGGGGGAAGTTCCATGGCATCACCAGGATCAC 280714
||| :|||
seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-103-840A-1
seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 441529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
alignment_scores:
Quality: 80.00 Length: 164
Ratio: 1.013 Gaps: 8
Percent Similarity: 48.171 Percent Identity: 23.171
alignment_block:
US-09-327-750D-12 x US-09-103-840A-1
Align seg 1/1 to: US-09-103-840A-1 from: 1 to: 441529
5 HisGlnGluAsnGluMetGluGlnProLeuGlnAsn...GlyGluG 20
||| :|||
280110 CATCCGCCGACGAATTCGCGAGAAATCGTCGTAATTCGCGCGTGA 280159
||| :|||
20 uAspArgProValGlyGlyGlyGlyGly.....HisGlnProA 33
||| :|||
280160 CGATGAACCTGCTGGCGCGATGCGAGATTGGCGCTGTTCTCCACCGG 280209
||| :|||
33 laGlyAsnAsnAsnAsnHis.....AsnHisAsnHisAsn 46
||| :|||
280210 CCGGTGACGGCGGTGCTGACCGCGGCTCCAGATCGGCGCGGACGATAAC 280259
||| :|||
47 HisHisArgArgGlyGlnAlaArg..... 54
||| :|||
280260 GATGAACGGTGGCTGCGCGCGAGCTCGAGCACGGTTCGCTTCGT 280309
||| :|||
55 .....ArgLeuAlaProAsnPheA 61
||| :|||
280310 TACCGCGCATAGCCGCCACCGATTGGCGCGGCTCGCTCCGCTCAGC 280359
||| :|||
61 rgTrpAlaIleProAsnArgGlnMetAsnAspGlyLeuGlyGly..... 75
||| :|||
280360 GTGGCCCGCGGACCGCGGATCAGCAGGATGCTTCGACGCGGTCGCGA 280409
||| :|||
76 .....AspGlyAs 78
||| :|||
280410 GCTAACAGCAACGCTCTGGAAGACGCGCTCGGGGAACCGCGCTCGGCGGA 280459
||| :|||
78 pAspMetGluMetPheMetGluGluMetArgGluIleArgArgLysLeuA 95
||| :|||
280573 TCAGCTCGGCGGATACAGCGCGCATTCGCGCACGTTTCGACGCGCTGTTG 280622
||| :|||
95 rgGluLeuGlnLeu.....ArgAsnCysLeu.....Arg 104
||||| :|||
280623 AGCAGCGGACGTTTCGGCGCATCAGTCCGCGTGGCGGAACCAACGCGC 280672
105 IleLeuMetGlyGluLeuSerAsnHisHis.....AspHis 116
||| :|||
280673 TTGCCACAGGGGGAAGTTCCATGGCATCACCAGGATCAC 280714
||| :|||
seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-103-840A-2
seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
alignment_scores:
Quality: 80.00 Length: 164
Ratio: 1.013 Gaps: 8
Percent Similarity: 48.171 Percent Identity: 23.171
alignment_block:
US-09-327-750D-12 x US-09-103-840A-2
Align seg 1/1 to: US-09-103-840A-2 from: 1 to: 4403765
5 HisGlnGluAsnGluMetGluGlnProLeuGlnAsn...GlyGluG 20
||| :|||
280223 CATCCGCCGACGAATTCGCGAGAAATCGTCGTAATTCGCGCGTGA 280272
||| :|||
20 uAspArgProValGlyGlyGlyGlyGly.....HisGlnProA 33
||| :|||
280273 CGATGAACCGTTCGCGCGATGCGAGATTGGCGCTGTTCTGCACCGG 280322
||| :|||
33 laGlyAsnAsnAsnAsnHis.....AsnHisAsnHisAsn 46
||| :|||
280323 CCGGTGACGGCGGTGCTGACCGCGGCTCCAGATCGGCGCGGACGATAAC 280372
||| :|||
47 HisHisArgArgGlyGlnAlaArg..... 54
||| :|||
280373 GATGAACGGTGGCTGCGCGCGAGCTCGAGCACGGTTCGCTTCGT 280422
||| :|||
55 .....ArgLeuAlaProAsnPheA 61
||| :|||
280423 TACCGCGCATAGCCGCCACCGATTGGCGCGGCTCGCTCCGCTCAGC 280472
||| :|||
61 rgTrpAlaIleProAsnArgGlnMetAsnAspGlyLeuGlyGly..... 75
||| :|||
280473 GTGGCCCGCGGACCGCGGATCAGCAGGATGCTTCGACGCGGTCGCGA 280522
||| :|||
76 .....AspGlyAs 78
||| :|||
280523 GCTAACAGCAACGCTCTGGAAGACGCGCTCGGGGAACCGCGCTCGGCGGA 280572
||| :|||
78 pAspMetGluMetPheMetGluGluMetArgGluIleArgArgLysLeuA 95
||| :|||
280573 TCAGCTCGGCGGATACAGCGCGCATTCGCGCACGTTTCGACGCGCTGTTG 280622
||| :|||

```

```

STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

```

```

alignment_scores:
  Quality: 81.00      Length: 144
  Ratio: 1.141      Gaps: 5
  Percent Similarity: 49.306      Percent Identity: 23.611

```

```
alignment_block:
```

```
US-09-327-750D-12 x US-09-036-987A-1/rev ..
```

```
Align seg 1/1 to reverse of: US-09-036-987A-1 from: 1 to: 80161
```

```

5 HisGlnGluAsnGluGluMetGluGlnProLeuGlnAsnGlyGluGluAs 21
||||| : : : : : : : : : : : : : : : : : : : : : : : :
69516 CACCAACTCGGTGATTCCAGAGCATCCACGACGACGCGCGGTCAACA 69467
21 pArgProValGlyGlyGlyGluGlyHisGlnProAlaGlyAsnAsn... 36
: : : : : : : : : : : : : : : : : : : : : : : :
69466 CCAATCGCGGACACCGCGTTCGGTCCACAGGTGGCGGCAACGCGCCC 69417
37 .....AsnAsnAsnHis.....Asn 42
69416 CCCAACAGGCCGTCGCCCGCGGAATCAGCACGGTTCGGTGGGATCCGC 69367
43 HisAsnHisAsnHis.....HisArgArgGlyG1 52
||||| : : : : : : : : : : : : : : : : : : : : : : : :
69366 CACAAACCGCGGACGTCGATCTCCCGGCGACCGGAAGCAAGCGGGCG 69317
52 nAlaArgArgLeuAlaPro.....AspPheArgTrpAlaIleP 65
: : : : : : : : : : : : : : : : : : : : : : : :
69316 CTCGCACCGCGCCTTCGCGAGGCGCAACTCCGTTCCATCGACCCCAAC 69267
65 roAsnArgGlnMetAsnAspGlyLeuGlyAspGlyAspMetGlu 81
: : : : : : : : : : : : : : : : : : : : : : : :
69266 GCACCTCGCAACGACGCGCATTCGCGGTTCCGTCGACGTCGCGCAG 69217
82 MetPheMetGluGluMetArgGluIleArgArgLysLeuArgGluLeu 98
: : : : : : : : : : : : : : : : : : : : : : : :
69216 CACAAACTGCCCGGGTCTCCACTCGCGGCGACGCAACACCCCAACA 69167
98 nLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHis. 114
| ||| ||| ||| : : : : : : : : : : : : : : : : :
69166 ACGCGCGCATGTGCCAGCTCTCATCCACCGTCACTCGGAGCGTCGTACC 69117
115 .....AspHisHisAspGluPheCysLeu 122
||||| : : : |||||

```

```

69116 GCTGCACACGTCACCAACCAACGACGTCGCTC 69085
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-370-700-1
seq_documentation_block:
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

```

```

alignment_scores:
  Quality: 81.00      Length: 144
  Ratio: 1.141      Gaps: 5
  Percent Similarity: 49.306      Percent Identity: 23.611

```

```
alignment_block:
```

```
US-09-327-750D-12 x US-09-370-700-1/rev ..
```

```
Align seg 1/1 to reverse of: US-09-370-700-1 from: 1 to: 80161
```

```

5 HisGlnGluAsnGluGluMetGluGlnProLeuGlnAsnGlyGluGluAs 21
||||| : : : : : : : : : : : : : : : : : : : : : : : :
69516 CACCAACTCGGTGATTCCAGAGCATCCACGACGACGCGCGGTCAACA 69467
21 pArgProValGlyGlyGlyGluGlyHisGlnProAlaGlyAsnAsn... 36
: : : : : : : : : : : : : : : : : : : : : : : :
69466 CCAATCGCGGACACCGCGTTCGGTCCACAGGTGGCGGCAACGCGCCC 69417
37 .....AsnAsnAsnHis.....Asn 42
69416 CCCAACAGGCCGTCGCCCGCGGAATCAGCACGGTTCGGTGGGATCCGC 69367
43 HisAsnHisAsnHis.....HisArgArgGlyG1 52
||||| : : : : : : : : : : : : : : : : : : : : : : : :
69366 CACAAACCGCGGACGTCGATCTCCCGGCGACCGGAAGCAAGCGGGCG 69317
52 nAlaArgArgLeuAlaPro.....AspPheArgTrpAlaIleP 65
: : : : : : : : : : : : : : : : : : : : : : : :
69316 CTCGCACCGCGCCTTCGCGAGGCGCAACTCCGTTCCATCGACCCCAAC 69267
65 roAsnArgGlnMetAsnAspGlyLeuGlyAspGlyAspMetGlu 81
: : : : : : : : : : : : : : : : : : : : : : : :
69266 GCACCTCGCAACGACGCGCATTCGCGGTTCCGTCGACGTCGCGCAG 69217
82 MetPheMetGluGluMetArgGluIleArgArgLysLeuArgGluLeu 98
: : : : : : : : : : : : : : : : : : : : : : : :
69216 CACAAACTGCCCGGGTCTCCACTCGCGGCGACGCAACACCCCAACA 69167
98 nLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHis. 114
| ||| ||| ||| : : : : : : : : : : : : : : : : :
69166 ACGCGCGCATGTGCCAGCTCTCATCCACCGTCACTCGGAGCGTCGTACC 69117
115 .....AspHisHisAspGluPheCysLeu 122
||||| : : : |||||

```

```

;
;
;
528 CGGCACCGCGCTG.....GAACATCGACCG 553
;
63 lalle.....ProAsnArg...GlnMetAsnAspGlyLeu...GlyGly 75
;
554 CAGTCCAGCGCCTCGTCGACCGTCGACGACGACGCTCTCGAAGCGGA 603
;
76 AspGlyAspAspMetGluMetPheMetGluGluMetArg..... 88
;
604 GATCAACGAGATCAAGAAGTCGGCGCGCGCAACGCGCGCTGCGGA 653
;
89 .....GluileArgArgGlyLeuA 95
;
654 TCGAGCGCACCGCGGGAAGTTCACCAAGCTGATCTCCGGCGGCGACGG 703
;
95 rgGluLeuGlnLeuArgAsnCysLeu...Arglle..... 105
;
704 ATCTACTCCAGCAGCGACGCTGCTCGCTCGGCTTCAACGTCGCGCAGCG 753
;
105 ..... 105
;
754 CAGCACCTACTACTCTGACCGCGCGCCACTGCACGCGGCGCGACCA 803
;
106 LeuMetGlyGluLeuSerAsnHisHisAsp...HisHisAsp 118
;
804 CCTGGTGGCGAAGTTCGCGCGCGCACCGTCTCGCGCACGAC 846
;
seq_name: /cqn2_6/ptodata/2/ina/6B_COMB.seq:US-09-171-969-1
seq_documentation_block:
; Sequence 1, Application US/09171969
; Patent No. 6284533
; GENERAL INFORMATION:
; APPLICANT: Thomas, Lawrence J.
; TITLE OF INVENTION: PLASMID-BASED VACCINE FOR TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 75 State Street, Suite 2300
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1807
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/171.969
; FILING DATE: 01 May 1997 (01.05.97)
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/640.713
; FILING DATE: 01 May 1996 (01.05.96)
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/802.967
; FILING DATE: 21 February 1997 (21.02.97)
; ATTORNEY/AGENT INFORMATION:
; NAME: Leon R. Yankwich
; REGISTRATION NUMBER: 30,237
; REFERENCE/DOCKET NUMBER: TCS 414.1 PCT (05872)
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1488 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL:
; ANTI-SENSE:
; FEATURE: Structural coding sequence for
; FEATURE: mature rabbit CDP
```

```

;
;
;
; NAME/KEY:
; LOCATION:
; PUBLICATION INFORMATION:
; AUTHORS: Nagashima, Mariko, et al.
; TITLE: Cloning and mRNA tissue distribution of rabbit cholesteryl ester trans
; JOURNAL: J. Lipid Res.
; VOLUME: 29
; ISSUE:
; PAGES: 1643 - 1649
; DATE: 1988
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1488
;
US-09-171-969-1
;
alignment_scores:
; Quality: 81.00 Length: 101
; Ratio: 1.373 Gaps: 3
; Percent Similarity: 58.416 Percent Identity: 28.713
;
alignment_block:
;
US-09-327-750D-12 x US-09-171-969-1 ..
;
Align seg 1/1 to: US-09-171-969-1 from: 1 to: 1488
;
5 HisGlnGluAsnGluGluMetGluGlnProLeuGlnAsnGlyGluGluAs 21
;
30 CATCGTGTGTCGCATCACCAGCCGCCCTCTT.....GGTGTGACC 73
;
21 pArgProValGlyGlyGluGlyHisGlnProAlaGlyAsnAsnAsnA 38
;
74 AAGAGACGGCCAAAGTGGTCCAGACGGCCCTTCACGCGCGCGGTATCCG 123
;
38 snAsnAsnHisAsnHisAsnHisAsnHisArgGlyGlnAlaArg 54
;
124 GAGCTCAGCGCGAGAGGGCGGTGATGCTCTCGCGCGGTCAAGTACGG 173
;
55 ArgLeuAlaProAsnPheArgTtpAlaIleProAsnArgGlnMetAsnA 71
;
174 GCTGCCACACCTCCAGATCAGCCACCTGTC....CATGCCAGCAGCAGG 220
;
71 pGlyLeuGlyGlyAspGlyAspMetGluMetPheMetGluGluMetA 88
;
221 TGGAGCTGGTGGAGCGCAAGACCATCGACGTGCCATCCAGACGTCGTC 270
;
88 rgGluIleArgArgLysLeuArgGluLeuGlnLeuArgAsnCysLeuA 104
;
271 GTGGTCTTCAAGGGGAC...CCTGAACACTACAGCTACACGAGTCCTGGGG 317
;
105 Ile 10'
;
318 GTT 32'
;
seq_name: /cqn2_6/ptodata/2/ina/6A_COMB.seq:US-09-036-987A-1
seq_documentation_block:
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J. J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
```

[illegible]

OM of: US-09-327-750D-12 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Mar 11, 2002 3:36 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=framet.p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101154\_20340/app\_query.fasta\_1.1472  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPEXT=4.500 -GAPEXT=0.050 -GAPEXT=10.000 -GAPEXT=0.500  
-GAPEXT=6.000 -GAPEXT=7.000 -GAPEXT=10.000 -GAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09327750.@CGL1\_1\_99 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT\_THREADS=1

## Search information block:

Query: US-09-327-750D-12  
Query length: 124  
Database: Issued\_Patents\_NA:\*  
Database sequences: 351203  
Database length: 113238999  
Search time (sec): 146.090000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-118-319-1 +	84.50	144.01	2.27	2353	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-118-319-1 +	83.00	144.66	2.08	1631	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-171-969-1 +	81.50	143.37	2.46	1373	Pate
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-036-987A-1 -	81.00	141.56	3.10	1488	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-370-700-1 -	81.00	101.63	519.83	80161	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2 +	80.00	59.49	1.1e-05	4403765	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1 +	80.00	59.48	1.1e-05	4411529	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-861-464-3 +	79.00	129.10	15.34	3455	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-396-001-3 +	79.00	129.10	15.34	3455	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-323-433A-3 +	79.00	129.10	15.34	3455	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-155-988-1 +	78.50	123.34	32.11	5552	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-023-173-4 +	77.00	136.18	6.19	1140	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-387-942C-18 +	77.00	136.05	6.29	1155	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-07-951-715A-6 +	77.00	127.46	18.92	2721	Pate
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-951-715A-6 +	77.00	124.59	27.33	3624	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-459-488A-6 +	77.00	124.59	27.33	3624	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-595A-6 +	77.00	124.59	27.33	3624	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-504B-6 +	77.00	124.59	27.33	3624	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-549-7 +	77.00	124.59	27.33	3624	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-053-549-7 +	77.00	124.59	27.33	3624	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-317-310A-1 +	77.00	123.39	31.91	4088	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-945-283-1 +	77.00	116.13	80.95	8438	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-387-942C-1 +	77.00	112.12	135.32	12588	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-789-275-1 +	76.50	127.83	18.05	2373	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-677-862-1 +	76.50	124.78	26.70	3218	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-252-571-1 +	76.50	124.78	26.70	3218	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-434-065-1 +	76.50	124.78	26.70	3218	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-245-295-5 -	76.50	120.57	45.82	4900	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-481-130-5 -	76.50	120.57	45.82	4900	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-656-984A-5 -	76.50	120.57	45.82	4900	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-485-604-5 -	76.50	120.57	45.82	4900	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-487-595-5 -	76.50	120.57	45.82	4900	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-863-790-26 -	76.50	120.57	45.82	4900	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-296-749-26 -	76.50	120.57	45.82	4900	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-245-295-8 -	76.50	120.21	47.95	5077	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-481-130-8 -	76.50	120.21	47.95	5077	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-656-984A-8 -	76.50	120.21	47.95	5077	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-485-604-8 -	76.50	120.21	47.95	5077	

/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-487-595-8 - 76.50 120.21 47.95 5077  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-024-020B-8 - 76.50 117.25 70.14 6826  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-356-952-12 - 76.50 98.66 76.11 43676  
/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-415-751-47 - 76.00 134.65 7.53 1086  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-025-691-2 + 76.00 132.90 9.42 1294

seq\_name: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:PCT-US92-06840-1

seq\_documentation\_block:  
; Sequence 1, Application PC/TUS9206840  
; GENERAL INFORMATION:  
; APPLICANT: Shi, Yang  
; APPLICANT: Seto, Edward  
; APPLICANT: Shenk, Thomas  
; TITLE OF INVENTION: Y11 TRANSCRIPTION FACTOR AND METHODS OF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
; STREET: 1180 Avenue of the Americas - 7th Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-8403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06840  
; FILING DATE: 19920814  
; CLASSIFICATION:  
; CLASSIFICATION: AU 1805  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/746,485  
; FILING DATE: 16-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dennis, Manette  
; REGISTRATION NUMBER: 30,623  
; REFERENCE/DOCKET NUMBER: M-12594 CIP (1570-8)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 382-0700  
; TELEFAX: (212) 382-0888  
; TELEX: 236925  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2353 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: Hela cells derived from cervical  
; TISSUE TYPE: carcinoma  
; CELL TYPE: tumor cells  
; CELL LINE: Hela  
; IMMEDIATE SOURCE:  
; LIBRARY: D98/AH-2  
; CLONE: p14-1 or pY11  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 241..1485  
; PCT-US92-06840-1

## alignment\_scores:

Quality: 84.50 Length: 86  
Ratio: 1.798 Gaps: 3  
Percent Similarity: 54.651 Percent Identity: 29.070

**THIS PAGE BLANK (USPTO)**

```
US-09-327-750D-13 x AAC10889  ..
Align seg 1/1 to: AAC10889 from: 1 to: 692

 3 AsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAsnGlyG1 19
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 AATGTC AACAGGAAATGATGATAAAGATCAAAAGGAGCAAGTGGCTAA 331
19 uGluAspArgProLeuGly.....GlyGlyGluGlyHisGlnP 32
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
332 TAAAGGGGAGCCCTTGGCCCTACCTTTGRTATGYTRGTGAATACGTGTGTC 381
32 roAlaGlyAsnArgArgGly...GlnAlaArgArgLeuAlaProAsnPhe 47
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
382 CTAGAGGAATCGTAGCGGTTCCGCGTTAGGCAGCCCATCTCCAGTAT 431
48 ArgTrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspG1 64
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
432 AGATGGGATATGATGCATAGCTTGGAGAACCCACAGGCAAGGATGANAGA 481
64 yAspAspMetCluIlePheMetGluGluMetArgGluIleArgArgLysL 81
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
482 AGAATATGGAAAGGATTGGGGAGGAGGTGACACAGCTGATGGAAAGC 531
81 euArgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeu 97
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
532 TGAGGGAAAGCAGTTGAGTCATAGTCTGCGGGCAGTCAGCAGCTGAC... 578
98 SerAsnHisHisAspHisAspGluPheCysLeuMetPro 111
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
579 CCCCTCACCATGACCATCATGATGAGTTTTCNNNATGCC 620
```





Align seg 1/1 to: AAF58581 from: 1 to: 862

```

3 AsnIleHisGlnGluAsn.....GluGluMetGluGlnPrometG1 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
271 AATGTCACACGAGAAATGATGAAAAAGATGAAAGAGGACCAAGTTGCTAA 320
16 nAsnGlyGluGlu...AspArgProLeuGlyGlyGlyHisGlnP 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
321 TAAAGGGAGCCCTTGGCCCTACCTTTGATGTTAGTGAATAGTGTGTCG 370
32 roLaGlyAsnArgArgGly...GlnAlaArgArgLeuAlaProAsnPhe 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
371 CTAGAGAAACCGTAGCGGTTCCGGCTTAGGCGACCCATCCTGCAGTAT 420
48 ArgTrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyGlyaspG1 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 AGATGGGACATATGATGATAGGCTTGGAGAGCCACAGGCAAGGATGAGAGA 470
64 yAspAspMetGluIlePheMetGluGluMetArgGluIleArgArgLysL 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
471 GGAGATATGGAAGGATGCGGAGGAGGTGACACAGCTGATGAAAGC 520
81 euArgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeu 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
521 TGAGGGAAAAGCAGTTGAGTCATAGTTTGGCGGCAGTCAGCACTGAT... 567
98 SerAsnHisHisAspHisAspGluPheCysLeuMetPro 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
568 CCCCCTCACCATGACCATCAGATGAGTGTGCTTATGCCCC 609

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF59611

seq\_documentation\_block:

ID AAF59611 standard; cDNA; 898 BP.

XX AAF59611;

XX 24-APR-2001 (first entry)

XX Human cell cycle and proliferation protein CCYPR-22 cDNA, SEQ ID NO:76.

XX Cell cycle and proliferation protein; CCYPR; human; agonist;  
 KW antagonist; gene therapy; detection; gene therapy;  
 KW transgenic animal disease model; immune disorder;  
 KW developmental disorder; cell signalling disorder;  
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;  
 KW menstrual cycle disorder; bacterial infection; ss.

OS Homo sapiens.

XX WO200107471-A2.

PN 01-FEB-2001.

XX 21-JUL-2000; 2000WO-US19948.

XX 21-JUL-1999; 99US-0145075.

PR 08-SEP-1999; 99US-0153129.

XX 10-NOV-1999; 99US-0164647.

PA (INCY-) INCYTE GENOMICS INC.

XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;

PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;

XX WPI; 2001-112727/12.

XX P-PSDB; AAB60474.

XX Human cell cycle and proliferation proteins and polynucleotides are

PT used to treat, diagnose and prevent immune, developmental and cell

PT signalling disorders and cell proliferative disorders including cancer -

XX

PS Claim 5; Page 181-182; 205pp; English.

XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human  
 CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.  
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
 CC associated with decreased expression of functional CCYPR, while CCYPR  
 CC antagonists are used to treat diseases or conditions associated with  
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
 CC that specifically bind to CCYPR, and in drug screening methods to  
 CC identify compounds that modulate the activity of CCYPR. CCYPR  
 CC nucleotides can be used to generate transgenic animal models of human  
 CC disease, and can be used in gene therapy in target cells with genetic  
 CC abnormalities with respect to the expression of CCYPR for the  
 CC treatment or prevention of a disorder associated with CCYPR.  
 CC Diseases which can be diagnosed, treated and prevented using CCYPR  
 CC proteins, nucleic acids, agonists or antagonists include immune,  
 CC developmental and cell signalling disorders, and cell proliferative  
 CC disorders including cancer. Specific examples of these disorders  
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
 CC diabetes mellitus, disorders of the menstrual cycle and infections  
 CC caused by bacteria.

XX Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;

alignment\_scores:

Quality: 175.00 Length: 114  
 Ratio: 2.465 Gaps: 4  
 Percent Similarity: 62.281 Percent Identity: 41.228

alignment\_block:

US-09-327-750D-13 x AAF59611 ..

Align seg 1/1 to: AAF59611 from: 1 to: 898

```

3 AsnIleHisGlnGluAsn.....GluGluMetGluGlnPrometG1 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 AATGTCACACGAGAAATGATGAAAAAGATGAAAGAGGACCAAGTTGCTAA 348
16 nAsnGlyGluGlu...AspArgProLeuGlyGlyGlyHisGlnP 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 TAAAGGGAGCCCTTGGCCCTACCTTTGATGTTAGTGAATAGTGTGTCG 398
32 roLaGlyAsnArgArgGly...GlnAlaArgArgLeuAlaProAsnPhe 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
399 CTAGAGGAAAACCGTAGCGGTTCCGGCTTAGGCGACCCATCCTGCAGTAT 448
48 ArgTrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyGlyaspG1 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
449 AGATGGGACATATGATGATAGGCTTGGAGAGCCACAGGCAAGGATGAGAGA 498
64 yAspAspMetGluIlePheMetGluGluMetArgGluIleArgArgLysL 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
499 GGAGATATGGAAGGATGCGGAGGAGGTGACACAGCTGATGGAAGC 548
81 euArgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeu 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
549 TGAGGGAAAAGCAGTTGAGTCATAGTCTGCGGCGCAGTCAGCACTGAT... 595
98 SerAsnHisHisAspHisAspGluPheCysLeuMetPro 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
596 CCCCCTCACCATGACCATCAGATGAGTGTGCTTATGCCCC 637

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AA03880

seq\_documentation\_block:

ID AA03880 standard; cDNA; 662 BP.

XX AA03880;

XX



PR	29-JAN-1999;	99US-0118078.
XX		
XX	(SEAR )	SEARLE & CO G D.
PA		
XX		
XX	Bunch RT,	Curtis SW, Rodi CP, Morris DL;
PI		
XX	WPI;	2000-505977/45.
DR		
XX	New nucleic acid encoding a carcinogenic biomarker, induced by	
PT	phenobarbital treatment of rat hepatocytes, useful for identifying	
PT	carcinogenic compounds -	
PT		
XX	Claim 1;	Page 73; 240pp; English.
XX		
CC	AAA87080 to AAA87656 represent nucleic acid sequences (N1) encoding a	
CC	carcinogenesis biomarkers. The carcinogenesis biomarkers are induced by	
CC	treating rat hepatocytes with phenobarbital. The nucleic acids are	
CC	useful for identifying carcinogenic compounds. The nucleic acid molecules	
CC	can be used to derive probes and/or primers for detecting or inducing	
CC	carcinogenesis, respectively.	
XX		
XX	Sequence 187 Bp;	39 A; 48 C; 40 G; 60 T; 0 other;
SQ		

```
alignment_scores:      Quality: 193.00      Length: 35
                       Ratio: 5.514      Gaps: 0
                       Percent Similarity: 100.000      Percent Identity: 100.000
```

```
alignment_block:
US-09-327-750D-13 x AAA87147
```

Align seg 1/1 to: AAA87147 from: 1 to: 187

```

77 IleArgArgLysLeuArgGluLeuGlnLeuArgAsnCysLeuArgIleLe 93
|||||
|||||
|||||
|||||
|||||
2 ATCCGGAGAAAGCTTAGGGAGCTGCAGTTCAGAAATGCTCGCTATTCT 51
|||||
|||||
|||||
|||||
|||||
93 uMetGlyCylLeuSerAsnHisHisAspHisHisAspGluPheCysLeuM 110
|||||
|||||
|||||
|||||
52 TATGGGGAGCTCTCTTAATCACCACGACCATCAGGATGAATTCCTGCCTTA 101
|||||
|||||
110 etPro 111
|||||
102 TGCCT 106

```

seq\_name: /SIDS2/qcdata/geneseq/geneseq/NA2001.DAT:AAH45143

seq\_documentation\_block:  
ID AAH45143 standard; cDNA; 792 BP.

DT 07-SEP-2001 (first entry)

AA	Human brain expressed X-linked protein, hBex, coding sequence.
DE	
XX	
KW	Human; brain expressed X-linked protein; cytostatic; auditory; nontropic;
KW	hBex; dysembryoplasia; hereditary disease; cancer; tumour; deafness;
KW	X-chromosome-binding mental retardation; lissencephalous disease; ss.

Id

Mao Y, Xie Y;

WPI; 2001-397944/42.  
P-PSDB: AAB99224.

Isolated human brain-expressed X-linked polypeptide used to diagnose and treat of dysembryoplasia, hereditary diseases, cancer, tumor, deafness and X-chromosome-binding mental retardation -

Claim 5: Page 22: 30pp: Chinese.

The present sequence is the coding sequence for a human brain-expressed x-linked protein (hBex). hBex and its coding sequence are useful in the diagnosis and treatment of dysembryoplasia, hereditary diseases, cancer, tumours, deafness, x-chromosome-binding mental retardation and lissencephalous disease. hBex is also useful for screening mimics, agonists, or inhibitors, and in peptide fingerprinting identification. hBex coding sequence can be used as primers or probes, or in producing gene chips or microarrays.

Sequence 792 BP; 214 A; 172 C; 219 G; 187 T; 0 other;

alignment_scores:		
Quality:	187.50	Length: 113
Ratio:	2.679	Gaps: 3
Percent Similarity:	61.947	Percent Identity: 42.478

```
alignment_block:
US-09-327-750D-13 x AAH45143
```

Align seq 1/1 to: AAH45143 from: 1 to: 792

```

1   MetalAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
    ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
212 ATGGAAAAATGCCAACCAAAAGTGAAGAAAAGGACGACTGTCTAATAA 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

17 nGlyGluGlu... AspArgProLeuGlyGlyGlyGluGlyHisGlnProA 33
    :||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
262 AGGGAGAGCCCTTGGCCCTCCCTTTGGATGCTGTGGTAATACTGTGCCTA 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

33 IaGlyAsnArgArgGly... GlnIaArgArgLeuAlaProAsnPheArg 48
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 GAGGAATCGTAGGCGGTTCGCGCTAGGCAGCCCATCCTGCAGTATAGA 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

49 TrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAs 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
362 TGGGATATCATCATAGGCTTTGGGAGACCACAGCAGGAGGATGAGAGAAGA 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

65 pAspMetGluIlePheMetGluGluMetArgGluIleArgArgLysLeua 82
    :|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
412 GAATATGGAAAGGATTGGGGAGGGGGTGACAGCTGATCGGAAAGCTGA 461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

82 rgGluLeuGluArgAsnCysLeuArgIleLeuMetGlyGluLeuSer 98
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
462 GGGAAAACAGTTGAGTTCATTAAGTCTGCGGGCAGTCAGCACTGAC...CCC 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

99 AsnHisHisAspIleHisAspGluPheCysLeuMetPro 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
509 CCTCACCATGACCATCATGATGAGTTTTTGCTTTATGCC 547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq name: /SIPDS2/qcadata/geneseq/geneseq/NA2001.DAT:AAH75810

seq\_documentation\_block:  
ID AAH75810 standard; cDNA: 767 BP;

17-OCT-2001 (first entry)

xx Human X chromosome linked gene expression protein 14 coding sequence  
DE  
xx  
KW Human; X chromosome linked gene expression protein 14; cancer;

286 AGAAGATCCGCCATTGTGGAGGGGTGAAGCCAGAGCCCTGGAGAA 335  
 35 sn...ArgArgGlyGlnAlaArgLeuAlaProAsnPheArgTrpAla 50  
 336 ATATCAGCGGGGGCGAGTTAGGCGACTTGCTCCTAATTTTCGATGGCC 385  
 51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67  
 386 ATACCTAATAGGCATATTGAGCAATGAAGCG.....AGAGATGATGT 429  
 67 tGluIlePheMetGluMetGluMetArgGluIleArgArgLysLeuArgGluL 84  
 430 AGAAGGTTTGTAGGCGAGATGATGGAATCAAGAGAAGACTAGGGAAC 479  
 84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGlyLeuSerAsnHis 100  
 480 AGCAGATGAGCGACTATATCGCGTTCTCAACTCTGAACTGACACCAT 529  
 101 HisAspHisAspGluPheCysLeuMetPro 111  
 530 TATGAC.....TTTGGCTCATACCT 550

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT: AAC85548

seq\_documentation\_block:

ID AAC85548 standard; cDNA; 1364 BP.

AC AAC85548;

DT 04-JUN-2001 (first entry)

DE cDNA encoding CDIFF-4, Incyte ID No. 1990956CB1.

KW Human; cell differentiation; CDIFF; agonist; antagonist; epilepsy;  
 cell proliferation; Alzheimer's disease; schizophrenia disorder;  
 arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.  
 KW Homo sapiens.

Key Location/Qualifiers  
 CDS 250..612  
 FT /\*tag= a  
 FT /product= "CDIFF-4"

XX WO200119860-A2.

XX 22-MAR-2001.

XX 14-SEP-2000; 2000WO-US25435.

XX 15-SEP-1999; 99US-0154140.

XX 06-DEC-1999; 99US-0169155.

XX (INCY-) INCYTE GENOMICS INC.

XX Tang YT, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;  
 PI Baughn MR, Lu DAM, Bandman O, Shih LL, Patterson C;

XX WPI; 2001-211447/21.

XX P-PSDB; AAB47126.

XX Isolated polypeptides and polynucleotides involved in cell  
 PT differentiation are used for treatment, prevention and diagnosis of  
 PT cell proliferative, developmental and neurological disorders e.g.  
 PT cancer and Alzheimer's disease -

XX Claim 5; Page 121; 137pp; English.

XX The sequences given in AAC85545-72 encode human polypeptides involved  
 CC in cell differentiation (CDIFF). CDIFF polypeptides and agonists of  
 CC these are used to treat a disease or condition associated with  
 CC decreased expression of functional CDIFF. An antagonist of CDIFF is  
 CC used to treat a disease or condition associated with over expression

CC of functional CDIFF. CDIFF polypeptides may be used for the treatment,  
 CC prevention and diagnosis of cell proliferative, developmental and  
 CC neurological disorders, such as Alzheimer's disease, schizophrenia  
 CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus  
 CC and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus  
 CC REX-3. This sequence maps to chromosome 1 within the interval from  
 CC 152.2 to 157.4 centimorgans, to chromosome 3 within the interval from  
 CC 157.4 to 158.0 centimorgans, and to the X chromosome within the interval  
 CC from 104.9 to 150.3 centimorgans.

XX Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;

alignment\_scores:

Quality: 277.00 Length: 111

Ratio: 3.298 Gaps: 4

Percent Similarity: 75.676 Percent Identity: 54.955

alignment\_block:

US-09-327-750D-13 x AAC85548 ..

Align seg 1/1 to: AAC85548 from: 1 to: 1364

3 AsnIleHisGlnGluAsnGluGluMetGluGln...PrometGlnAsnG 18

||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

295 AATGCCCAACAAGAAACGAAGAGGAGGAGGCCGCCACGCGAGAAATGA 344

18 yGluGluAspArgProLeuGlyGlyGlyGlyGlyHisGlnProAlaGlyA 35

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

345 AGAAGAAATCCGCCATTGTGGAGGGGTGAAGGCCAGAGCTGGAGAA 394

35 sn...ArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50

||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

395 ATATCAGCGGGGGCGAGTTAGGCGACTTGTCCTAATTTTCGATGGGCC 444

51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

445 ATACCTAATAGGCATATTGAGCAATGAAGCG.....AGAGATGATGT 488

67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84

:||| :|||:|||||:|||||:|||||:|||||:|||||:|||||

489 AGAAGGTTTGTAGGCGCAGATGATGGAATCAAGAGAAGACTAGGGAAC 538

84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGlyLeuSerAsnHis 100

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

539 AGCAGATGAGCGCACTATATGCGCTTCCAAACTCTCTGAACCTGACACCAT 588

101 HisAspHisAspGluPheCysLeuMetPro 111

:|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

589 TATGAC.....TTTGGCTCATACCT 609

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT: AAA87147

seq\_documentation\_block:

ID AAA87147 standard; DNA; 187 BP.

XX AC AAA87147;

XX 08-JAN-2001 (first entry)

XX Rat hepatocyte carcinogenesis biomarker nucleic acid SEQ ID NO:71.

XX Rat; phenobarbital; carcinogenesis marker; carcinogenesis; detection;  
 KW identification; carcinogenic; probe; primer; ds.

XX Rattus norvegicus.

XX WO200044902-A2.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US00503.

XX



|||||  
369 ATACCAATAGGCAGATCAATGATGGGTGAGATGATAT 418  
67 tGluilePheMetGluGluMetArgGluileArgArgLysLeuArgGluL 84  
|||||  
419 GGAATATTCATGGAGAGATGAGAGAAATCAGAAGAAATCTTAGGGAGC 468  
84 euGlnLeuArgAsnCysLeuArgTleLeuMetGlyGluLeuSerAsnHis 100  
|||||  
469 TGCATTTGAGGAATGCTCGGTATCCTTATGGGGGAKC1CTCTAATCAC 518  
101 HisAspHisHis 104  
|||||  
519 CATGACCATCAT 530

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH03517

seq\_documentation\_block:  
ID AAC01004 standard; cDNA; 485 BP.

AC AAC01004;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 1002.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG00998.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 1002; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.

XX Sequence 485 BP; 125 A; 113 C; 159 G; 84 T; 4 other;

alignment\_scores:  
Quality: 521.00 Length: 99  
Ratio: 5.427 Gaps: 0

Percent Similarity: 96.970 Percent Identity: 96.970  
alignment\_block:  
US-09-327-750D-13 x AAC01004 ..  
Align seg 1/1 to: AAC01004 from: 1 to: 485  
6 GlnGluAsnGluGluMetGluGluProMetGlnAsnGlyGluGluAspAr 22  
|||||  
187 CAGAAACCAAGAGATGGAGCAGCCCTATGCAGATGGAGAGGAGACCG 236  
22 gProLeuGlyGlyGlyGluGlyHisGlnProAlaGlyAsnArgArgGlyG 39  
|||||  
237 CCCTTTGGAGGAGGTGAAGGCCACAGCCTGCAGAAATCGACGGGAC 286  
39 InAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnArgGln 55  
|||||  
287 AGGCTCGCCGACTTGCCTTAATTTTCGATGGGCATACCCAATAGGCAG 336  
56 IleAsnAspGlyMetGlyGlyAspGlyAspMetGluIlePheMetG1 72  
|||||  
337 ATCAATGATGGGTGGGTGRAGATGGAGATGATATGGAATATTATCGGA 386  
72 uGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArgAsnC 89  
|||||  
387 GGAGATCAGAGAAATCAGACAAACTTAGGAGCTGCAKTTGAGGAATT 436  
89 ysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisHis 104  
|||||  
437 GTCTGCGTATCCTTATGGGGAKCTCTCTAATCACCATGACCATCAT 483

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH03517

seq\_documentation\_block:

ID AAH03517 standard; cDNA; 865 BP.

XX AAH03517;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:352.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs -

XX Claim 1; SEQ ID 352; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602





|||||  
412 GAAATCGAGGGGACGCTCGCGACTGCCCCCTAAATTTTCGATGGGCC 461  
51 IleProAsnArgInIleAsnAspGlyMetClyGlyAspGlyAspPme 67  
|||||  
462 ATACCCAAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGATGAT 511  
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84  
|||||  
512 GGAATATTTCATGAGGAGATGAGAGAAATCAGAGAAACTTAGGAGGC 561  
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100  
|||||  
562 TGCAGTTGAGGAATGTCTGGTATCTTATGGGGAGCTCTCTAATCAC 611  
101 HisAspHisHisAspGluPheCysLeuMetPro 111  
|||||  
612 CATGACCATGATGATGAAATTTTGCCCTTATGCCT 644

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21748

seq\_documentation\_block:

ID: AAF21748 standard; DNA; 917 BP.

AC AAF21748;

XX 27-MAR-2001 (first entry)

XX Human breast and ovarian cancer associated antigen gene SEQ ID 135.  
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
KW neurotropic; neurprotection; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antinflammatory; antiulcer; vulnerary; anticonvulsant;  
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;  
KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
KW cardiovascular disorder; wound healing; neurological disease; ds.

XX Homo sapiens.

XX WO200055173-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05881.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-611515/58.

XX P-PSDB; AAB58845.

PT New human breast and ovarian cancer associated gene sequences and the  
PT polypeptides encoded by these genes, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -

XX Claim 1; Page 581-582; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are  
XX associated with breast and ovarian cancer. Included in the invention are  
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
XX isolation and characterisation of the DNA and protein sequences of the  
XX invention. The breast and ovarian cancer associated DNA, protein, agonist  
XX or antagonist sequences exhibit cytostatic; immunosuppressive;  
XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
XX antidiabetic; antinflammatory; antiulcer; vulnerary; anticonvulsant;  
XX antibacterial; antifungal; antiparasitic and cardiant activity. The

CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
CC particularly breast and ovarian cancer. The nucleic acid sequences,  
CC proteins, agonists and agonists may also be used in the diagnosis,  
CC prevention and treatment of immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemias; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases.

XX Sequence 917 BP; 228 A; 203 C; 270 G; 211 T; 5 other;

alignment\_scores:

Quality: 582.00 Length: 106  
Ratio: 5.491 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-13 x AAF21748 ..

Align seg 1/1 to: AAF21748 from: 1 to: 917

6 GlnGluAsnGluGluMetGluGlnProMetGlnAsnGlyGluGluAspAr 22  
|||||

306 CAGCAAAACGAAAGAGATGGAGCAGCCTATGCAGAAATGGAGAGACCG 355  
|||||

22 gProLeuGlyClyGlyGluGlyHisGlnProAlaGlyAsnArgArgGlyG 39  
|||||

356 CCCCTTTGGGAGGAGGTGAAGGCCACCAGCCTGCAGAAATCGAGGGGAC 405  
|||||

39 InAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnArgGln 55  
|||||

406 AGGCTCCCGACACTTGCCCTTAATTTTCATGGGCCATACCCCAATAGGCG 455  
|||||

56 IleAsnAspGlyMetGlyGlyAspGlyAspMetGluIlePheMetGl 72  
|||||

456 ATCAATGATGGGATGGGTGGAGATGGAGATGATGGAATATTTCATGGA 505  
|||||

72 uGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArgAsnC 89  
|||||

506 GGAGATGAGAGAAATCAGAGAAACTTAGGGAGCTGCAGTTGAGGAATT 555  
|||||

89 ysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisHisAsp 105  
|||||

556 GTCTGCGTATCTTATGGGGAGGCTCTTAATCACCATGACCATCATCAT 605  
|||||

106 GluPheCysLeuMetPro 111  
|||||

606 GAATTTTGCCCTTATGCCT 623

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF23528

seq\_documentation\_block:

ID: AAF23528 standard; DNA; 700 BP.

XX AC AAF23528;

XX DT 22-MAR-2001 (first entry)

XX Mouse NADE DNA.

XX Neurotrophin receptor; p75-NTR; NCF-induced apoptosis;  
KW neurogenetic disease; NF-kappaB; ds.

XX Mus sp.

XX WO200075278-A2.

XX 14-DEC-2000.

XX 07-JUN-2000; 2000WO-US15621.

XX

OM of: US-09-327-750D-13 to: N\_Geneseq\_1101:\* out\_format : pfs  
Date: Mar 11, 2002 3:42 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+ p2n\_model -DEV=xlh  
-Q=Cgcn2\_1/USPTO\_spool/US09327750/runat\_11032002\_101154\_20362/app\_query.fasta\_1.1472  
-DB=N\_Geneseq\_1101 -Qfmt=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsm62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09327750\_CGcn1\_1.330 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-13

Query length: 111

Database: N\_Geneseq\_1101:\*

Database sequences: 930621

Database length: 428662619

Search time (sec): 355.560000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF23529 +		609.00	1187.76	4.4e-58	891
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAF23529 +		582.00	1134.20	4.2e-55	917
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF23528 +		564.50	1102.28	2.5e-53	700
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAF23528 +		537.00	1050.66	1.9e-50	532
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AA01005 +		519.00	1019.98	9.8e-49	485
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA01004 +		521.00	532.88	1.3e-21	865
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AA03517 +		277.00	529.49	2.0e-21	1229
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH13750 +		277.00	528.48	2.3e-21	1364
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH85548 +		193.00	381.92	3.4e-13	187
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH45143 +		187.50	357.12	8.2e-12	792
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH5810 +		175.00	332.76	1.9e-10	767
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH58581 +		175.00	331.63	2.1e-10	862
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH59611 +		175.00	331.23	2.3e-10	898
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA03880 +		169.00	322.34	7.1e-10	662
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA03879 +		169.00	321.91	7.4e-10	692
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA03879 +		169.00	321.83	7.5e-10	698
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH82586 +		168.00	324.63	5.3e-10	426
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH60367 +		162.00	306.02	5.7e-09	858
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA01003 +		108.00	206.81	0.0019	401
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH72409 +		105.00	190.32	0.0159	1197
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH14082 +		102.50	196.05	0.0076	397
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH135463 +		102.50	196.05	0.0076	397
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH13935 +		102.50	196.05	0.0076	397
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH121436 +		91.50	177.55	0.0819	285
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH16725 +		91.50	177.55	0.0819	285
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH17131 +		91.50	177.55	0.0819	285
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH68557 +		91.50	163.90	0.4716	1170
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH157960 +		91.50	163.26	0.5119	1250
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA098961 +		91.50	162.81	0.5421	1309
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH159746 +		91.50	161.17	0.6889	1551
/SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:AAQ63293 +		88.50	135.01	19.17	12588
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH81469 +		87.50	117.80	174.20	60873
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH21610 +		87.50	100.90	1.5e+03	349980
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH81490 +		87.50	87.24	8.8e+03	1437668
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH19956 +		84.50	151.53	2.30	1007
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ34631 +		82.50	128.35	45.02	7365
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:AAH37051 +		82.00	132.67	25.88	4255
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:AAH44922 +		80.00	146.02	4.67	711
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH87356 +		80.00	140.62	9.34	1243
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH64101 +		80.00	140.62	9.34	1243

/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AA04966 + 80.00 140.62 9.34 1243  
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH87357 + 80.00 140.40 9.61 127  
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA064099 + 80.00 140.40 9.61 127  
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AA04964 + 80.00 140.40 9.61 127  
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH1328 + 79.00 149.09 3.15 422

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF23529

seq\_documentation\_block:

ID AAF23529 standard; DNA; 891 BP.

XX

AC AAF23529;

DT 22-MAR-2001 (first entry)

XX

DE Human NADE DNA.

KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;

KW neurogenetic disease; NF-kappaB; ds.

XX

OS Homo sapiens.

XX

PN WO200075278-A2.

PD 14-DEC-2000.

PF 07-JUN-2000; 2000WO-US15621.

XX

PR 07-JUN-1999; 9905-0327750.

XX

PA (UYCO ) UNIV COLUMBIA NEW YORK.

XX

PI Sato T;

XX

DR WPI; 2001-061707/07.

XX

PT New p75-neurotrophin receptor-associated cell death executor (NADE) and the gene encoding NADE, useful for modulating the activity of p75NTR and for detecting neurodegenerative diseases -

PT

PS Disclosure: Fig 1; 134pp; English.

XX The present invention relates to a purified polypeptide capable of binding neurotrophin receptor (p75-NTR). The invention is useful for binding and modulating the activity of p75NTR. The peptide mediates NGF-induced apoptosis, which plays an important role in neurogenetic diseases. The peptide of the invention and p75NTR are useful for inhibiting NF-kappaB activation in a cell or a subject, for inducing caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase and fragment nuclear DNA in a cell by co-expression of (1) and p75-NTR.

XX

SQ Sequence 891 BP; 251 A; 182 C; 224 G; 234 T; 0 other;

alignment\_scores:

Quality: 609.00 Length: 111

Ratio: 5.486 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-13 x AAF23529

Align seg 1/1 to: AAF23529 from: 1 to: 891

1 MetAlaAsnIleHisGlnGluAsnGluGluMetGluGlnProMetGlnAs 17

|||||

312 ATGCCAAATATTACCAGGAAACGAAGAGATGAGAGCGCTATGCAGAA 361

|||||

17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAlaG 34

|||||

362 TGGAGAGGAGGAGCGCCCTTTGGAGAGGAGTGAGGCCACCGCTGCAG 411

|||||

34 LyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTyrAla 50

**THIS PAGE BLANK (USPTO)**





```

21  pargProLeuGlyGlyGlyGluGlyHisGlnProAlaGlyAsnArgG 38
2168  ACGCGTGTGGGGCAGCGAACAATCCACCATCCAGGAGGCGAACGACGTG 2217
38  lyGlnAla.....ArgArgLeuAlaPro 45
2218  TTCAAGGAGAACTACGTGACCTGCCCGCGCACCTTCAACGAGTGCACCC 2267
46  AsnPhe.....ArgTrpAl 50
2268  CACCTACCTGTACCAGAAGATCGCGGAGAGCGAGCTGAAGGCCTACACCC 2317
50  atleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyGly 65
2318  GCTACCAGCTGCGCGGCTACATCGAGGACAGCAGGACCTGGAGATCTAC 2367
56  .....AspMet 67
2368  CTGATCCGCTACACGCCCAAGCAGCAGAGACCCCTGGAGCTGCCGCGCACCGA 2417
68  GluIlePheMetGluGluMetArgGluIleArgGlyLysLeuArgGluLe 84
2418  GAGCCTGTGGCCCTTGAGGCTGGAGAGCCCATCGGCGCCTGCGGCGAGC 2467
84  uGln...LeuArgAsnCysLeuArgIle..... 92
2468  CCAACCGCTGCGCCCCCCTTCGAGTGGAGTGAACCCCGCACCTGGACTGCAGC 2517
93  ..LeuMet.GlyGluLeuSerAsnHisHisAspHisHis 104
2518  TGCCCGCAGCGCGAGAAGTGCGCCACCCACAGCCACCAC 2556

seq_name: /cgm2_6/ptodata/2/ins/6A_COMB.seq:us-08-459-504B-6

seq_documentation_block:
; Sequence 6, Application US/08459504B
; Patent No. 6075185
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttle, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6075185artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,504B
; FILING DATE:
; CLASSIFICATION:

```



```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Other
US-09-179-558-62

alignment_scores:
    Quality: 75.50      Length: 103
    Ratio: 1.452        Gaps: 5
    Percent similarity: 50.485    Percent identity: 25.243

alignment_block:
US-09-327-750D-13 x US-09-179-558-62/rev ..

Align seg 1/1 to reverse of: US-09-179-558-62 from: 1 to: 936

3 AsnIleHISgInGluAsnGluGluMetGluGln.....Pr 14
      ::::::::::::::: :::: ::::
730 AGCCGTGCATCATGACCCCGGGTTCCAAACAGCTCCCGCAACACCACCA 681

14 oMetGlnAsnGlyGluGluAspArgProLeuGlyGlyGluGlyHisG 31
      ::::::::::::::: :::: ::::
680 GGTACGCGAGGAGAGATGAGGCCCATGTGGCGCGGTGATGGACAGC 631

31 lnProAlaGlyAsnArgArgGlyGlnAlaArgLeuAlaProAsnPhe 47
      ::::: ::::: ::::: :::::
630 ACGGCCACGGCAGGCCCGGGGACAACCTGCTGGCA..... 590

48 ArgTrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAsp 64
      ::::: ::::: ::::: :::::
589 .....GAGTCTCTCGCGTTCGGCCGAGGAGGTGGAGC 558

64 yAspAspMetGluIlePheMetGluGluMetArgGluIleArgArgLys 81
      ::::: ::::: ::::: :::::
557 CATCGATCAGGATGTG.....CTGCACGCGCTTCTCTCAAGC 520

81 euArgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeu 97
      ::::: ::::: ::::: :::::
519 TC.....CTTAGTAAATGCATCCAG..... 500

98 SerAsnHis 100
      :::::
499 ...AACCAC 494

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-951-715A-6

seq_documentation_block:
; Sequence 6, Application US/07951715A
; Patent No. 5625136
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne

```



2069 CGAGCCCTCCGCGCGCGCGCGCGCTTC.....CACCAACGCGC 2112  
 49 pAlaIleProAsnArgGlnIleAsn.....AspGlyMetGlyC 62  
 2113 CCACCTCCGCGCGCGCGCGCGCACCTCCGCGCGCGGATCCGCGCGGAG 2162  
 62 lyAspGlyAspMetGlu 68  
 2163 GAGAGGAGAGGAGGAGGAG 2182

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-179-558-64

```
seq_documentation_block:
; Sequence 64, Application US/09179558
; Patent No 6180612
; GENERAL INFORMATION:
; APPLICANT: Hockensmith, Joel W.
; APPLICANT: Muthuswami, Rohini
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TARGETING DNA METABOLIC PROCESSES USING
; TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; STREET: PENNIE & EDMONDS LLP
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/179,558
; FILING DATE: 27-OCT-1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 09/060,470
; FILING DATE: 15-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 60/063,898
; FILING DATE: 31-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9426-005-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 892 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other
US-09-179-558-64
```

alignment\_scores:  
 Quality: 75.50 Length: 103  
 Ratio: 1.452 Gaps: 5  
 Percent similarity: 50.485 Percent Identity: 25.243

alignment\_block:

US-09-327-750D-13 x US-09-179-558-64/rev ..

Align seg 1/1 to reverse of: US-09-179-558-64 from: 1 to: 892

3 AsnIleHisGlnGluAsnGluGluMetGluGln.....Pr 14

```
686 AGCTGCATCAGCACCCCGGGTTCCAAAACAGCTCCCCGAACACCA 637
14 oMetGlnAsnGlyGluGluAspArgProLeuGlyGlyGlyGlyHisG 31
636 GGTACGCCGAGGAGAAGATGAGGCCCATGTTGGCGCGGTGATGACAGC 587
31 lnProAlaGlyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsn 47
586 ACGGCCACGCGCGCGCGGGAACAAGTAACTGCTGTGCA..... 546
48 ArgTrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAsp 64
545 .....GAGGTCTCTCGCGGTGCGCGGAGGAGTGAGC 514
64 yAspAspMetGluIlePheMetGluGluMetArgGluIleArgArgLys 81
513 CATCGATACGATGTG.....CTGCACGCGCTTCTCTCAAGC 476
81 euArgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeu 97
475 TC.....CTTAGTAATTGCATCCAG..... 456
98 SerAsnHis 100
455 ...AACCAC 450
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-179-558-62
seq_documentation_block:
; Sequence 62, Application US/09179558
; Patent No 6180612
; GENERAL INFORMATION:
; APPLICANT: Hockensmith, Joel W.
; APPLICANT: Muthuswami, Rohini
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TARGETING DNA METABOLIC PROCESSES USING
; TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; STREET: PENNIE & EDMONDS LLP
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/179,558
; FILING DATE: 27-OCT-1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 09/060,470
; FILING DATE: 15-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 60/063,898
; FILING DATE: 31-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9426-005-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 936 base pairs
; TYPE: nucleic acid
```





**3137** TGGAGACGGGGAATGAGCAGCATCCGGAGGAGATGAAGGAGGCAGCGCTAC 3186  
:  
**41** ArgArgLeuAlaIleProAsnPheArgTrpAlaIle..... 51  
::: |||:::|||||||  
**3187** CATGCCACCACCCCACAAGTACAGACGGGCTGGTTTATGCTCCTCCGCA 3236  
:  
**52** .ProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyLysAspMetG 68  
:|||:::|||||  
**3237** TCCGGATCAAGAGCGCGCCTCCGGATCGGGATCGGAT..... 3273  
:  
**68** luIlePheMetGluMetArgGluIleAerArgLysLeuArgGluLeu 84  
:|||:::||||| : :: |  
**3274** ..ATCATGTGGATCGCGCTACAATTTCGAGTGACAAGTCCAAGGAGCTC 3321  
:  
**85** GlnLeuArgAsnCysLeuArgIle..... LeuMetGlyGI 96  
:::::::::::  
**3322** AACATGCAGCGCAACATACGATGTCAGTCCCCAGCACGACATGATGCCCA 3371  
:  
**96** uLeuSerAsnHisHisAspHisHis 104  
: ||| |||||  
**3372** CTATTCCGCGCATCATCGSCACCAT 3396

```
61 GlyGlyAspGlyAspMetGluLeuPheMetGluGluMetArgGlu.. 76
||||| ||| :|||: |||:|
216 GACGGTGCATCAAGATGCTGCAGCGCTGCAGCGCAGAACGTCACCGGCATGG 265
77 .....IleArgArgLysLeuArgGluLeuGln 85
:|||||: :|||:|
266 TGCCTCGGCTACGGCGAGGAAACCACTTCGG 302
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-387-942C-1
seq_documentation_block:
; Sequence 1, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: ERTESVAG, HELGA
; APPLICANT: VALLA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,942C
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1809-106P
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12588 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Azotobacter vinelandii
; STRAIN: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 290..1951
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2227..6438
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6702..9695
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9973..12588
; US-08-387-942C-1
```

```
alignment_scores:
Quality: 88.50
Ratio: 2.011
Length: 79
Gaps: 4
```

```
Percent similarity: 55.696 Percent Identity: 36.709
alignment_block:
US-09-327-750D-13 x US-08-387-942C-1
Align seg 1/1 to: US-08-387-942C-1 from: 1 to: 12588
18 GlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAlaG1 34
|||||:| ||||| |||||:|
2298 GCGGATCGACGCGCGCCACCGCGCGCGCGCGCGCGCTACCTGCCG 2347
34 yAsnArg...ArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpA 50
||||| ||||| ||||| |||||
2348 CCGCGGAATATCGGGTCAGCGCGCGGAGGAGCC.....TTCCGATGGT 2391
50 laile.....ProAsnArgGlnIleAsnAspGlyMet 60
||||| ||||| ||||| |||||
2392 TGCTGTGACCATCAAGACACGTCATATCTCGCGCGCGGATGGCGGA 2441
61 GlyGlyAspGlyAspMetGluIlePheMetGluGluMetArgGlu.. 76
||||| ||||| ||||| |||||
2442 GACGGTGCATCAAGATGCTGCAGCGCTGCAGCGCAGAACGTCACCGCATGG 2491
77 .....IleArgArgLysLeuArgGluLeuGln 85
:|||||: :|||:|
2492 TGCCTCGGCTACGGCGAGGAAACCACTTCGG 2528
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US96-02331-12
seq_documentation_block:
; Sequence 12, Application PC/TUS9602331
; GENERAL INFORMATION:
; APPLICANT: The Board of Trustees of the Leland Stanford Junior
; APPLICANT: University
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: Methods and Compositions for Altering
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/02331
; FILING DATE: 09-FEB-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,495
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 8600-0153.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```



553 TATATGGCTTCCAACTCTGAACTGACACCACTATGAC..... 594

118 pGluPheCysLeuMetPro 124

|||||.....

595 .....TTTGGCTCATACCT 609

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAA87147

seq\_documentation\_block:

ID AAA87147 standard; DNA; 187 BP.

XX

XX AAA87147;

XX 08-JAN-2001 (first entry)

XX Rat hepatocyte carcinogenesis biomarker nucleic acid SEQ ID NO:71.

XX Rat; phenobarbital; carcinogenesis marker; carcinogenesis; detection;

KW Identification; carcinogenic; probe; primer; ds.

XX Rattus norvegicus.

XX WO200044902-A2.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US00503.

XX 29-JAN-1999; 99US-0118078.

XX (SEAR ) SEARLE & CO G D.

XX Bunch RT, Curtis SW, Rodi CP, Morris DL;

XX WPI; 2000-505977/45.

XX New nucleic acid encoding a carcinogenic biomarker, induced by

PT phenobarbital treatment of rat hepatocytes, useful for identifying

PT carcinogenic compounds -

XX Claim 1; Page 73; 240pp; English.

XX AAA87080 to AAA87656 represent nucleic acid sequences (N1) encoding a

CC carcinogenesis biomarkers. The carcinogenesis biomarkers are induced by

CC treating rat hepatocytes with phenobarbital. The nucleic acids are

CC useful for identifying carcinogenic compounds. The nucleic acid molecules

CC can be used to derive probes and/or primers for detecting or inducing

CC carcinogenesis, respectively.

XX Sequence 187 BP; 39 A; 48 C; 40 G; 60 T; 0 other;

alignment\_scores:

Quality: 193.00

Ratio: 5.514

Percent Similarity: 100.000

Length: 35

Gaps: 0

Percent Identity: 100.000

alignment\_block:

US-09-327-750D-12 x AAA87147

Align seg 1/1 to: AAA87147 from: 1 to: 187

90 IleArgArgLysLeuArgGluLeuGlnLeuArgAsnCysLeuArgIleLe 106

|||||.....

2 ATCCGGAGAAGCTTAGGGAGCTGCAGTTGAGAAATTGCTGCTATTCT 51

|||||.....

106 uMetGlyGluSerAsnHisHisAspHisHisAspGluPheCysLeu 123

|||||.....

52 TATGGGGAGCTCTCTAATCACCAGCACCATCAGATGAATTCCTGCTTA 101

|||||.....

123 etPro 124

|||||.....

102 TGCCT 106

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH45143

seq\_documentation\_block:

ID AAH45143 standard; cDNA; 792 BP.

XX

XX AAH45143;

XX 07-SEP-2001 (first entry)

XX Human brain expressed X-linked protein, hBex, coding sequence.

XX Human; brain expressed X-linked protein; cytostatic; auditory; nontropic;

KW hBex; dysembryoplasia; hereditary disease; cancer; tumour; deafness;

KW X-chromosome-binding mental retardation; lissencephalous disease; ss.

XX Homo sapiens.

XX WO200140286-A1.

XX 07-JUN-2001.

XX 27-NOV-2000; 2000WO-CN00502.

XX 30-NOV-1999; 99CN-0124179.

XX (BIOR-) BIOROAD GENE DEV LTD SHANGHAI.

XX Mao Y, Xie Y;

XX WPI; 2001-397944/42.

XX P-PSDB; AAB99224.

XX Isolated human brain-expressed X-linked polypeptide used to diagnose

PT and treat of dysembryoplasia, hereditary diseases, cancer, tumor,

PT deafness and X-chromosome-binding mental retardation -

XX Claim 5; Page 22; 30pp; Chinese.

XX The present sequence is the coding sequence for a human brain-expressed

CC X-linked protein (hBex). hBex and its coding sequence are useful in the

CC diagnosis and treatment of dysembryoplasia, hereditary diseases, cancer,

CC tumours, deafness, X-chromosome-binding mental retardation and

CC lissencephalous disease. hBex is also useful for screening mimics,

CC agonists, or inhibitors, and in peptide fingerprinting identification.

CC hBex coding sequence can be used as primers or probes, or in producing

CC gene chips or microarrays.

XX Sequence 792 BP; 214 A; 172 C; 219 G; 187 T; 0 other;

alignment\_scores:

Quality: 184.00

Ratio: 2.453

Percent Similarity: 57.252

Length: 131

Gaps: 5

Percent Identity: 36.641

alignment\_block:

US-09-327-750D-12 x AAH45143

Align seg 1/1 to: AAH45143 from: 1 to: 792

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17

|||||.....

212 ATGGAATGCCAACCAAGAAATGAAGAAAGAGCAGCTTCCTAATAA 261

|||||.....

17 nGlyGluGlu...AspArgProValGlyGlyGlyGlyHisGlnProA 33

|||||.....

262 AGGGAGCCCTTGGCCCTCCCTTTGGATGCTGTGTAATACTGTGCTTA 311

|||||.....

33 laGlyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisHisArg 49

|||||.....

312 GAGGAAT.....CGTAGG 325

```

50  ArgGlyClnAlaArgArgLeuAlaProAsnPheArgTrpAlaIle..... 64
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326  CGGTTCCGGCTTAGCAGAGCCCATCTCGCAGTATAGATGGGATATGATGCA 375
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
65  .....ProAsnArgGlnMetAsnAspGlyLeuGlyGlyAspG 77
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
376  TAGGCTTGAGAACACACAGCGCATGAGAA..... 409
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77  lyAspAspMetGluMetPheMetGluLeuMetArgGluIleLeuArgArgLys 93
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410  .GAGAAATATGGAAGAGGATTGGGGAGGGGGTGAGACAGCTCATCGGAAAAG 457
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
94  LeuArgGluLeuClnLeuArgAsnCysLeuArgIleLeuMetGlyGluLe 110
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
458  CTGAGGGAAAAACGATTGAGTCATAGCTCGCGGCACCTAGCACTGAC... 505
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110  uSerAsnHisHisAspHisHisAspGluPheCysLeuMetPro 124
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
506  .CCCCCTCACCATGACCATCATGATGACTTTTGCTTTATGCC 547
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq\_name: /SIDS2/gcgcdata/geneseq/geneseq/NA2001.DAT:AAH75810

seq\_documentation\_block:  
ID AAH75810 standard; cDNA; 767 bp.

AA  
AC  
AAH75810;

DT 17-OCT-2001 (first entry)

XX DE · Human X chromosome linked gene expression protein 14 coding sequence.

Human; X chromosome linked gene expression protein 14; cancer;  
KW  
HIV infection; cytostatic; anti-HIV; chromosome X; ss.  
KW

YY Homo sapiens.

XX  
PN  
CN1296969-A.XX  
PD  
30-MAY-2001.

AA  
PF 23-NOV-1999; 99CN-0124078.

XX  
PR 23-NOV-1999: 99CN-0124078.

PA (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.

PI Mao Y, Xie Y;

XX  
DR WPI: 2001-483897/53XX  
DR P-PSDB; AAG66407.

PT Polypeptide-human X chromosome Linked gene expression protein 14 and polynucleotide for coding said polypeptide -

XX  
PS  
claim 6: page 24 (Disclosure): 31pp: Chinese

CC The present sequence is the coding sequence for human X chromosome linked  
CC gene expression protein 14. The protein and coding sequence are useful  
CC for treating diseases e.g. cancer and HIV infection.  
AX

Sequence 767 BP; 224 A; 143 C; 215 G; 185 T; 0 other;

alignment_scores:		
Quality:	171.50	Length: 132
Ratio:	2.287	Gaps: 6
Percent Similarity:	56.818	Percent Identity: 36.364

alignment\_block:

US-09-327-750D-12 x AAH75810

Align seg 1/1 to: AAH75810 from: 1 to: 767

```

3  AsnValHisGlnGluAsn.....GluGluMetGluGlnProLeuG1 16
      : : : : : : : : : : : : : : : : : : : : : : : :
146 AATGTC AACCCAGGAAAATGATGAA AAAAGATGAAAAGGACGAAGTTGCTAA 195
      : : : : : : : : : : : : : : : : : : : : : : : :
16  nAsnGlyGluGlu...AspArgProValGlyGlyGlyGluGlyHisGlnP 32
      : : : : : : : : : : : : : : : : : : : : : : : :
196 TAAAGGGAGGACCCCTTGCCCTACCTTTTGATGCTTACTGATCTACTGTGTGC 245
      : : : : : : : : : : : : : : : : : : : : : : : :
32  roAlaGlyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHis 48
      : : : : : : : : : : : : : : : : : : : : : : : :
246 CTAGAGGAAAC.....CGT 259
      : : : : : : : : : : : : : : : : : : : : : : : :
49  ArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIle... 64
      : : : : : : : : : : : : : : : : : : : : : : : :
260 AGCGCGTTCCCGGTTCAGGACGCCCATCTGCAGTATAGATGGACATAAT 309
      : : : : : : : : : : : : : : : : : : : : : : : :
65  .....ProAsnArgGlnMetAsnAspGlyLeuGlyGlyA 76
      : : : : : : : : : : : : : : : : : : : : : : : :
310 GCATAGCGTTGGAGAGCCACAGGCAAGGATGAGACAG..... 346
      : : : : : : : : : : : : : : : : : : : : : : : :
76  spGlyAspMetGluMetPheMetGluGluMetArgGluIleArgArg 92
      : : : : : : : : : : : : : : : : : : : : : : : :
347 .....GAGAATATGGAAGGATTGGGGAGGAGGTGAGACAGCTGATGGAA 391
      : : : : : : : : : : : : : : : : : : : : : : : :
93  LysLeuArgGluLeuGluArgAsnCysLeuArgIleLeuMetGlyG1 109
      : : : : : : : : : : : : : : : : : : : : : : : :
392 AACCTGAGGGGAAAAGCAGTTGATGTCATAGTCTGCGGGCAGTGCAGCAGCTGA 441
      : : : : : : : : : : : : : : : : : : : : : : : :
109 uLeuSerAsnHisHisAspHisHisAspGluPheCysLeuMetPro 124
      : : : : : : : : : : : : : : : : : : : : : : : :
442 T...CCCCCTCACCATCATCAGTATGAGTTGGCTTATGCCCC 484
      : : : : : : : : : : : : : : : : : : : : : : : :

```

seq\_name: /SIDS2/qcdata/geneseq/geneseq/NA2001.DAT:AAI58581

seq\_documentation\_block:

ID AAI58581 standard; cDNA; 862 BP.

AC AAI58581;

22-OCT-2001 (first entry)

XX  
DE Human polynucleotide SEQ ID NO 784.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemolactatic;  
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
leukaemia; ss.

XX Homo sapiens.

XX  
PN WO200153312-A1XX  
PD  
26-JUL-2001XX  
PF 26-DEC-2000: 2000WQ-IIS34263

XX  
BB 21-JAN-2000: 2000US-0488725

PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042

PR 19-JUL-2000; 2000US-0620312.  
DB 03-AUG-2000; 2000US-0652450

PR 14-SEP-2000; 2000US-0662191.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
xx





alignment\_block:  
US-09-327-750D-12 x AAF59611 ..  
Align seg 1/1 to: AAF59611 from: 1 to: 898  
3 AsnValHisGlnGluAsn.....GluGluMetGluGlnProLeuG1 16  
|||||:|||||: ||||| |||||  
299 AATGTCACACAGGAATGAAAGATGAAAAGGAGCAAGTTGCTAA 348  
16 nAsnGlyGluGlu...AspArgProValGlyGlyGluGlyHisGlnP 32  
:|||||: |||||: |||||: |||||  
349 TAAAGGGGAGCCCTTGCCCTTACCTTTCAATGTTAGTGAATACTGTGTC 398  
32 roAlaGlyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHis 48  
|| |||||  
399 CTAGAGGAAC.....CGT 412  
49 ArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIle.. 64  
||||| :|||: |||||  
413 AGCGGTTCCGCGTTAGGCAGCCCATCTGCAGTATAGATGGGACATAAT 462  
65 .....ProAsnArgGlnMetAsnAspGlyLeuGlyGlyA 76  
:|||||: |||||: |||||: |||||  
463 GCATAGGCTTGGAGAGCCACAGGCAAGGATGAGAG..... 499  
76 spGlyAspAspMetGluMetPheMetGluGluMetArgGluIleArgArg 92  
:|||||: |||||: |||||: |||||: |||||: |||||  
500 .....GAGATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAA 544  
93 LysLeuArgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyG1 109  
|||||: |||||: |||||: |||||: |||||: |||||  
545 AAGCTGAGGGAAGGAGGATTGAGTCATAGTCTGCGGCGAGTGCAGCACTGA 594  
109 uLeuSerAsnHisHisAspHisHisAspGluPheCysLeuMetPro 124  
: |||||: |||||: |||||: |||||: |||||: |||||  
595 T...CCCCCTCACCATCACCATCAGATGAGTTTGGCTTATGCC 637  
seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH82586  
seq\_documentation\_block:  
ID AAH82586 standard; cDNA; 426 BP.  
XX  
AC AAH82586;  
XX  
DT 25-SEP-2001 (first entry)  
XX  
DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:210.  
XX  
KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;  
KW immunogenic; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200151513-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 16-JAN-2001; 2001WO-US01575.  
XX  
PR 14-JAN-2000; 2000US-0176722.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Algate PA;  
XX  
DR WPI; 2001-425866/45.  
XX  
PT Novel ovarian tumor proteins, and nucleic acids encoding them, used to  
PT treat and diagnose cancers, particularly ovarian cancer -  
XX  
PS Claim 5; Page 100; 338pp; English.  
XX  
CC AAH82377 to AAH83878 represent human ovarian tumour-associated

CC polynucleotide sequences which encode ovarian tumour proteins. The  
CC ovarian tumour protein and polynucleotide sequences have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. The  
CC ovarian tumour proteins and polynucleotides can be used to inhibit  
CC the development of cancer, particularly ovarian cancer. They can also  
XX be used to diagnose the onset and progression of cancer.  
SQ Sequence 426 BP; 86 A; 106 C; 90 G; 137 T; 7 other;

alignment\_scores:  
Quality: 168.00 Length: 31  
Ratio: 5.600 Gaps: 0  
Percent Similarity: 96.774 Percent Identity: 96.774  
alignment\_block:  
US-09-327-750D-12 x AAH82586 ..  
Align seg 1/1 to: AAH82586 from: 1 to: 426  
94 LeuArgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLe 110  
|||||: |||||: |||||: |||||: |||||: |||||  
1 CTTAGGGAGCTGCATTTGAGGAATTGCTGCGTATCCTTATGGGGAGCT 50  
110 uSerAsnHisHisAspHisHisAspGluPheCysLeuMetPro 124  
|||||: |||||: |||||: |||||: |||||: |||||  
51 CTCATATCACCATGACCATCATCATGAATTTGCCCTTATGCGCT 93

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH03880  
seq\_documentation\_block:  
ID AAC03880 standard; cDNA; 662 BP.  
XX  
AC AAC03880;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein 5' EST, SEQ ID NO: 3878.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
PA (GEST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
DR WPI; 2000-500381/45.  
DR P-PSDB; AAG03874.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 1; SEQ ID 3878; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been

CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
SQ Sequence 662 BP; 177 A; 144 C; 192 G; 137 T; 12 other;

## alignment\_scores:

Quality: 162.50 Length: 126  
Ratio: 2.355 Gaps: 4  
Percent Similarity: 54.762 Percent Identity: 35.714

## alignment\_block:

US-09-327-750D-12 x AAC03880 ..

Align seg 1/1 to: AAC03880 from: 1 to: 662

```
3 AsnValHisGlnGluAsn.....GluGluMetGluGlnProLeuG1 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 AATGTCACACGAGAAATGATGAAGAAGATGAAGAAGACCAAGTGCCTAA 301
16 nAsnGlyGluGlu...AspArgProValGlyGlyGlyGlyHisGlnP 32
:::|||||:|||||:|||||:|||||:
302 TAAAGGGGAGCCCTTGGCCCTACCTTTGRTGTYTRGTGAATACTGTGTGC 351
32 rolaGlyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHis 48
|| |||||
352 CTAGAGGAAT.....CGT 365
49 ArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIlePr 65
|||||:|||||:|||||:|||||:|||||:
366 AGCGGTTCCGCGTTAGGAGCCCTCCTGCAGTATAGATGGATATGAT 415
65 oAsnArgGlnMetAsnAspGlyLeuGlyGlyGlyAspGlyAspMetGluM 82
:::|||||:|||||:|||||:
416 GCATAGGCTTGGAGAACCAAGCAGGAGGATGANAGAAGAGAAATATGAAA 465
82 etPheMetGluGluMetArgGluIleArgArgLysLeuArgGluLeuGln 98
::: |||||:|||||:|||||:|||||:|||||:
466 GGATTGGGGAGGAGGTGAGACAGCTGATGGAAGCTGAGGGAAGAGCAG 515
99 LeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAs 115
||| ::: |||||:|||||:|||||:
516 TTGATCATAGTCTGGGGCAGTCAGCACTGAC...CCCCCTCACCATGA 562
115 pHisHisAspGluPheCysLeuMetPro 124
|||||:|||||:|||||:|||||:
563 CCATCATGATGAGCTTTTGCNNWATGCC 590
```

**THIS PAGE BLANK (USPTO)**

OM of: US-09-327-750D-12 to: GenEmbl:\* out\_format : pfs  
Date: Mar 11, 2002 3:33 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2.1/uspto.spool/US09327750/runat\_11032002.101154.20319/app\_query.fasta\_1.1472  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -MODALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -DOCS=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09327750@cgn2.1.8673  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIME=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-327-750D-12  
Query length: 124  
Database: GenEmbl\*  
Database sequences: 1472140  
Database length: -341344837  
Search time (sec): 4557.230000

## score\_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_ro:AF187066	+	688.00	1064.22	5.3e-51	700	AF187066 Mus musculus p75NTR-as
gb_ro:AF097440	+	688.00	1062.65	6.5e-51	854	AF097440 Mus musculus brain exp
gb_ro:AF187065	+	643.00	997.06	2.9e-47	519	AF187065 Rattus norvegicus p75N
gb_pr:BC003190	+	576.50	890.99	2.4e-41	793	BC003190 Homo sapiens, p75NTR-a
gb_pr:HUMOGC	+	576.50	890.07	2.7e-41	891	M38188 Human unknown protein fr
gb_pr:AF187064	+	576.50	890.07	2.7e-41	891	AF187064 Homo sapiens p75NTR-as
gb_sts:G72708	-	411.00	640.31	2.2e-27	421	G72708 MARC 4953-4954-991939031
gb_pr:HSV351F8	+	282.50	404.88	2.8e-14	45678	Z70719 Human DNA sequence frd
gb_pr:AK000959	+	284.50	405.55	2.6e-14	1229	AK000959 Homo sapiens cDNA FLJ
gb_pat:AX100231	+	264.50	404.73	2.9e-14	1364	AX100231 Sequence 32 from Pat
gb_pr:HS635913	+	264.50	373.75	1.5e-12	69648	AL035494 Human DNA sequence f
gb_pr:HS1714B7	+	240.50	333.97	2.5e-10	98274	Z99755 Human DNA sequence frd
gb_sts:G35294	-	235.50	368.21	3.1e-12	477	G35294 human STS SHGC-37409, se
gb_pr:AF183416	+	191.00	297.12	2.8e-08	642	AF183416 Homo sapiens ovarian g
gb_pr:AF220189	+	191.00	295.12	3.7e-08	828	AF220189 Homo sapiens uncharact
gb_pr:HS198P4	+	184.00	284.66	1.4e-07	35714	AL008708 Human DNA sequence f
gb_pr:AF097439	+	172.00	266.19	1.5e-06	785	AF097439 Mus musculus brain exp
gb_pr:AF251053	+	171.50	265.84	1.6e-06	744	AF251053 Homo sapiens X-linked
gb_pat:AX078272	+	171.50	264.35	1.9e-06	898	AX078272 Sequence 76 from Pat
gb_pr:HSV670H8	+	171.50	236.38	6.9e-05	31321	Z70233 Human DNA sequence frd
gb_pr:AL133348	-	171.50	234.34	8.9e-05	40584	AL133348 Homo sapiens DNA sequence
gb_pr:HSU080B1	+	171.50	234.25	9.0e-05	41029	AL022169 Homo sapiens chromos
gb_ro:AF097438	+	156.00	240.98	3.8e-05	835	AF097438 Mus musculus brain exp
gb_ro:AF051347	+	156.00	240.95	3.8e-05	838	AF051347 Mus musculus REX-3 mRN
gb_ro:AF097437	+	156.00	233.11	0.0001	2269	AF097437 Mus musculus Bex1 pro
gb_pr:HSU10564	+	132.50	204.36	0.0042	43952	Z92846 Human DNA sequence frd
gb_pr:AL357046	+	138.00	172.06	0.2626	154577	AL357046 Homo sapiens chrom
gb_htg:AC009282	-	138.00	172.04	0.2633	154959	AC009282 Homo sapiens clone
gb_pr:AL353305	+	138.00	170.63	0.3153	185257	AL353305 Human DNA sequence
gb_sts:G24641	-	122.50	193.21	0.0174	504	G24641 human STS WI-11354, sequ
gb_vi:AF181716	-	112.00	166.40	0.5425	1933	AF181716 Cynomolgus Epstein-Ba
gb_pr:HS112K5	-	106.50	126.08	95.54	109891	Z85987 Human DNA sequence fr
gb_htg:AL591766	+	106.00	121.35	175.26	181658	AL591766 Homo sapiens chrom
gb_htg:PFMAL4P3	+	104.00	113.78	462.96	321003	AL035476 Plasmodium falcipar
gb_htg:AC091781	-	103.00	117.62	282.95	162045	AC091781 Trypanosoma brucei
gb_in:LCU75355	+	101.00	145.11	8.32	3336	U75355 Lucilia cuprina ecyste
gb_vi:CH045963	-	100.50	148.62	5.30	1936	U45963 Cercopithecine herpesvi
gb_htg:AC079989	-	100.50	112.68	532.71	185650	AC079989 Rattus norvegicus d
gb_htg:AC079378	-	100.50	112.58	539.96	188150	AC079378 Rattus norvegicus d

gb\_ba:AE004899 - 100.00 134.31 33.27 10807 ! AE004899 Pseudomonas aerug  
gb\_pl:AB016876 + 100.00 123.16 138.96 4499 ! AB016876 Arabidopsis thali  
gb\_pl:NCB1D1 - 99.00 117.39 291.18 76072 ! AL355927 Neurospora crassa  
gb\_htg:AC011771 + 99.00 111.60 612.42 158807 ! AC011771 Homo sapiens chr  
gb\_pl:SCFPSIG + 97.00 140.33 15.37 2794 ! X54157 S.cerevisiae FPS1 (1

seq\_name: gb\_ro:AF187066

seq\_documentation\_block:

LOCUS AF187066 700 bp mRNA ROD  
DEFINITION Mus musculus p75NTR-associated cell death executor (Nade) mRNA,  
complete cds.

ACCESSION AF187066

VERSION AF187066.1 GI:8452897

KEYWORDS

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 700)

AUTHORS Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,

Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.

NADE, a p75NTR-associated cell death executor, is involved in

signal transduction mediated by the common neurotrophin receptor

p75NTR

J. Biol. Chem. 275 (23), 17566-17570 (2000)

MEDLINE 20298829

REFERENCE 2 (bases 1 to 700)

AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,

Hanaoka,T., Li,Y., Irie,S. and Sato,T.

Direct Submission

Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia

University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA

LOCATION/Qualifiers

1. .700

/organism="Mus musculus"

/strain="BALB/c"

/db\_xref="taxon:10090"

1. .700

/gene="Nade"

177. .551

/gene="Nade"

/note="Involved in the common neurotrophin receptor

p75NTR-mediated signal transduction; NADE"

/codon\_start=1

/product="p75NTR-associated cell death executor"

/protein\_id="AAF75131.1"

/db\_xref="GI:8452898"

/translation="MANYHQENEEQPLQNGQEPVGGGGHQPAAANNNNNNHNNH

HNHRRGQARRLAPNFWAIPNRMNDGLGGGPGDMENFMEMREIRKRLRELQRLNC

LRILMGLSNHHDHDEFLMP"

BASE COUNT 178 a 187 c 203 g 132 t

ORIGIN

alignment\_scores:

Quality: 688.00 Length: 124

Ratio: 5.548 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.387

alignment\_block:

US-09-327-750D-12 x AF187066

Align seg 1/1 to: AF187066 from: 1 to: 700

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17

177 ATGCCCAATCTCCACCAAGAAAGAGATGGAGCAGCCCTGCAGAA 226

17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlyHisGlnProAlaG 34

227 TGGACAGGAAGACGCCCTCTGGGAGGAGGTGAGGCGCCAGCCTGCTG 276

34 lYAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArgArg 50  
:|||||  
277 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 326  
51 GlyGlnAlaArgArgLeuAlaProAsnProAsnPheArgTrpAlaIleProAsnAr 67  
|||||  
327 GCCCAGGCTGCCGACTTGGCCCTTAACCTTCGATGGCCATTCCTCCACAG 376  
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84  
|||||  
377 GCAGATGAATGACGGTGGGTGGAGATGGAGATGCATGGAAATGTTC 426  
84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100  
|||||  
427 TGGAGGAGATGAGAGATCCGGAGAAGCTTAGGGAGCTACACTGAGA 476  
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH 117  
|||||  
477 AATTGTCTACGCATCTTATGGGGAGCTGTCTAACCAACCACGATCACCA 526  
117 sAspGluPheCysLeuMetPro 124  
:|||||  
527 TGATGAATTCTGCCTTATGCCT 548

seq\_name: gb\_ro:AF097440

seq\_documentation\_block:  
LOCUS AF097440 854 bp mRNA ROD 13-APR-1999  
DEFINITION Mus musculus brain expressed X-linked protein 3 (Bex3) mRNA,  
complete cds.  
ACCESSION AF097440  
VERSION AF097440.1 GI:4580593  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 854)  
Brown, A.L. and Kay, G.F.  
Bex1, a gene with increased expression in parthenogenetic embryos,  
is a member of a novel gene family on the mouse X chromosome  
Hum. Mol. Genet. 8 (4), 611-619 (1999)  
99172070  
Erratum: [[published erratum appears in Hum Mol Genet 1999  
May; 8(5):943]]  
2 (bases 1 to 854)  
Brown, A.L. and Kay, G.F.  
Direct Submission  
Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of  
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia  
Location/Qualifiers  
1. .854  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/db\_xref="dbEST:AA272375"  
/chromosome="X"  
/map="near Plp"  
/cissus\_type="pooled organs"  
1. .854  
/gene="Bex3"  
172. .546  
/gene="Bex3"  
/codon\_start=1  
/product="brain expressed X-linked protein 3"  
/protein\_id="AAD24431.1"  
/db\_xref="GI:4580594"  
/translation="MANVHQNEEMEQPLQGOEDRPVGGEGHQPAAANNHNNHNNH  
HNHRRGOARLAPNFWAIPNRQNDGLGGDDMEMEEMREIRKRLRELQLRNC  
LRLLGELSNHHDDHDECLMP"  
237 a 212 c 228 g 177 t

FEATURES

Source

gene

CDS

BASE COUNT

ORIGIN

alignment\_scores:  
Quality: 688.00 Length: 124  
Ratio: 5.548 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.387  
alignment\_block:  
US-09-327-750D-12 x AF097440 ..  
Align seg 1/1 to: AF097440 from: 1 to: 854  
1 MetAlaAsnValHisGlnGluAsnGluGluMetGluGlnProLeuGlnAs 17  
|||||  
172 ATGCCCAATGTCCACAGAAACGAAGACATGGAGCGCCCTGCCAGAA 221  
17 nGlyGluGluAspArgProValGlyGlyGlyGlyHisGlnProAlaG 34  
|||||  
222 TGGACAGGAAGACCGCCCTGTGGAGGAGGTGAGGGCCACGAGCTGCTG 271  
34 lYAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArgArg 50  
:|||||  
272 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 321  
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67  
|||||  
322 GCCCAGGCTGCCGACTTGGCCCTTAACCTTCGATGGCCATTCCTCCACAG 371  
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84  
|||||  
372 GCAGATGAATGACGGTGGGTGGAGATGGAGATGCATGGAAATGTTC 421  
84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100  
|||||  
422 TGGAGGAGATGAGAGATCCGGAGAAGCTTAGGGAGCTACAGCTGAGA 471  
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH 117  
|||||  
472 AATTGTCTACGCATCTTATGGGGAGCTGTCTAACCAACCACGATCACCA 521  
117 sAspGluPheCysLeuMetPro 124  
:|||||  
522 TGATGAATTCTGCCTTATGCCT 543

seq\_name: gb\_ro:AF187065

seq\_documentation\_block:

LOCUS AF187065 519 bp mRNA ROD 11-JUN-2000

DEFINITION Rattus norvegicus p75NTR-associated cell death executor (Nade)

mrna, complete cds.

ACCESSION AF187065

VERSION AF187065.1 GI:8452895

KEYWORDS

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 519)

Mukai, J., Hachiya, T., Shoji-Hoshino, S., Kimura, M.T., Nadano, D.,

Suvanto, P., Hanaoka, T., Li, Y., Irie, S., Greene, L.A. and Sato, T.A.

Nade, a p75NTR-associated cell death executor, is involved in

signal transduction mediated by the common neurotrophin receptor

p75NTR

J. Biol. Chem. 275 (23), 17566-17570 (2000)

MEDLINE 20298829

REFERENCE 2 (bases 1 to 519)

Mukai, J., Hachiya, T., Hoshino, S., Kimura, M., Nadano, D.,

Hanaoka, T., Li, Y., Irie, S. and Sato, T.

Direct Submission

TITLE Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia

JOURNAL University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA

FEATURES Location/Qualifiers

source 1. .519

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1. 519
/gene="Nade"
44..406
/gene="Nade"
/feature="involved in the common neurotrophin receptor
p75NTR-mediated signal transduction; NADE"
/codon_start=1
/product="p75NTR-associated cell death executor"
/protein_id="AAF75130.1"
/db_xref="GI:8452896"
/translation="MEQLQNGEDRPVGGEGHQPAGNNNNHHNNHHNNHHH
RRGQARRLAPNFWAIPNRMQNDGDDMEMEEMREIRKRLRELQRLNCLRL
MGELSNHHHDEFCLMP"
BASE COUNT 142 a 130 c 136 g 111 t
ORIGIN

alignment_scores:
  Quality: 643.00 Length: 125
  Ratio: 5.403 Gaps: 1
Percent Similarity: 95.200 Percent Identity: 92.800

alignment_block:
US-09-327-750D-12 x AF187065 ..

Align seg 1/1 to: AF187065 from: 1 to: 519

6 GlnGluAsnGluMetGluGlnProLeuGlnAsnGlyGluGluAspAr 22
|||||
29 CAGGAAACGAAGAAATGAGCAGCCCTGCAGAAATGGACAGGACCG 78
|||||
22 gProValGlyGlyGlyGlyHisGlnProAlaGlyAsnAsn.... 37
|||||
79 CCCCCTGGGAGGAGGTGAGGGCCACCGCTGCTGCAACACCAACACC 128
|||||
38 .....AsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisAsn 49
|||||
129 ACACCATATACACACACACACACACACACACACACATCATCACCG 178
|||||
50 ArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIlePro 66
|||||
179 AGAGGACAGCCCGCGACTTGCCTTAACCTTCGATGGCCATTCACCA 228
|||||
66 nArgGlnMetAsnAspGlyLeuGlyAspGlyAspMetGluMetP 83
|||||
229 CAGGCAGATGAATGATGGGTGGGTGGGATGGAGATGATATGGAATG 278
|||||
83 heMetGluGluMetArgGluIleArgArgLysLeuArgGluLeuGln 99
|||||
279 TCATGAGGAGATGAGAGATGATCCGAGAAAGCTTAGGAGCTGCAGTT 328
|||||
100 ArgAsnCysLeuArgIleLeuMetGlyGlyLeuSerAsnHisAspH 116
|||||
329 AGAAATGTCTGCTATCTTATGGGGAGCTCTCTAATACACCAACCA 378
|||||
116 sHisAspGluPheCysLeuMetPro 124
|||||
379 TCACGATGAATTCGCTTATGCT 403

seq_name: gb_pr:BC003190

seq_documentation_block:
LOCUS BC003190 793 bp mRNA PRI 12-JUL-2001
DEFINITION Homo sapiens, p75NTR-associated cell death executor; ovarian
granulosa cell protein (13kD), clone MGC:802 IMAGE:3357965, mRNA,
complete cds.
ACCESSION BC003190
VERSION BC003190.1 GI:13112030
KEYWORDS MGC.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 793)
REFERENCE Strausberg,R.
AUTHORS Direct Submission
TITLE Submitted (13-FEB-2001) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcysc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 6 Row: k Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7657043.

FEATURES             Location/Qualifiers
     source
       1..793
         /organism="Homo sapiens"
         /db_xref="LocusID:27018"
         /db_xref="taxon:9606"
         /clone="MGC:802 IMAGE:3357965"
         /tissue_type="Eye, retinoblastoma"
         /clone_lib="NIH_MGC_16"
         /lab_host="DH10B-R"
         /note="Vector: pOTB7"
         196..531
           CDS
             /codon_start=1
             /product="p75NTR-associated cell death executor; ovarian
granulosa cell protein (13kD)"
             /protein_id="AAH03190.1"
             /db_xref="GI:13112031"
             /translation="MANTHQNEEMQPMQNGEDRPLGGEGHQPAGNNRRRQARRLA
PNFWAIPNROIINDGMDGDDMEIFMEEMREIRKRLRELQRLNCLRLMGLSNHHH
HHDEFCLMP"
BASE COUNT 225 a 170 c 205 g 193 t
ORIGIN

alignment_scores:
  Quality: 576.50 Length: 124
  Ratio: 5.194 Gaps: 1
Percent Similarity: 89.516 Percent Identity: 84.677

alignment_block:
US-09-327-750D-12 x BC003190 ..

Align seg 1/1 to: BC003190 from: 1 to: 793

1 MetAlaAsnValHisGlnGluAsnGluGluMetGluGlnProLeuGlnAs 17
|||||
196 ATGCCAAATATTCACCAAGAAACGAGAGATGAGCAGGCTATGCAGAA 245
|||||
17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlyHisGlnProAlaG 34
|||||
246 TGGAGAGGAAGACCGCCCTTTGGGAGGAGGTGAAGGCCACACGCTGCAG 295
|||||

```

```

34  lyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArgArg 50
|||||
296  GAAAT.....CGACGG 306

51  GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67
|||||
307  GGACAGGCTGCGCGACTTGGCCCTTAATTTTCGATGGCCCATACCAANTAG 356

67  gGlnMetAsnAspGlyLeuGlyArgGlyAspGlyAspMetGluMetPheM 84
|||||
357  GCAGATCAATGATGGATGGGTGGAGATGGAGATGATGATAATATTCATCA 406

84  etGluGluMetArgGluLeuArgGlyLysLeuArgGluLeuGlnLeuArg 100
|||||
407  TGGAGGAGATGAGAGAAATCAGAAGAACTTGGGAGCTGCAGTTGAGG 456

101  AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH 117
|||||
457  AATTGTCTGCGTATCCTTATGGGGAGCTCTCTAATCACCATGACCATCA 506

117  sAspGluPheCysLeuMetPro 124
|||||
507  TGATGAATTTTGCCCTTATGCCT 528

```

seq\_name: gb\_pr:HUMOGC

```

seq_documentation_block:
LOCUS HUMOGC 891 bp mRNA 07-MAR-1995
DEFINITION Human unknown protein from clone PHGR74 mRNA, complete cds.
ACCESSION M38188 X56942
VERSION M38188.1 GI:189378
KEYWORDS
SOURCE Human ovarian granulosa cell line, cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Rapp,G., Freudenstein,J., Klaudiny,J., Mucha,J., Wempe,F.,
Zimmer,M. and Scheit,K.H.
TITLE Characterization of three abundant mRNAs from human ovarian
granulosa cells
JOURNAL DNA Cell Biol. 9 (7), 479-485 (1990)
MEDLINE 91025550
COMMENT [Draft entry and computer-readable sequence for [DNA 9, 479-485
(1990)]] kindly submitted
by K.H.Scheit, 27-AUG-1990.

```

FEATURES

```

source
    Location/Qualifiers
        1..891
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="PHGR74"
            /cell_type="granulosa"
            /tissue_type="ovary"
            <1..891
            /note="protein of unknown function"
            312..547
            /note="protein of unknown function"
            /codon_start=1
            /protein_id="AAA63232.1"
            /db_xref="GI:189379"
            /translation="MANIHOENEEMOPMONGEEDRPLGGEGHQPAGNRREGQARRLA
            PNFRWAINRINDMGDDGDDMEIFMEEMREIRKRLRELQRLNCLRLMGELSNHHD
            HHDFCLMP"
            361..534
            /note="protein of unknown function"
            /codon_start=1
            /protein_id="AAA63233.1"
            /db_xref="GI:189380"
            /translation="MERKALTALVEEVKATSLQEIDGDLADLPILFDGPPYPIGRSMGW
            VEMEMINKYSWRR"

```

CDS

```

JOURNAL J. Biol. Chem. 275 (23), 17566-17570 (2000)
MEDLINE 20298829
REFERENCE
AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
Hanaoka,T., Li,Y., Irie,S. and Sato,T.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES
    source
        1..891

```

```

alignment_scores:
    Quality: 576.50      Length: 124
    Ratio: 5.194        Gaps: 1
    Percent Similarity: 89.516      Percent Identity: 84.677

alignment_block:
US-09-327-750D-12 x HUMOGC ..
Align seg 1/1 to: HUMOGC from: 1 to: 891

1  MetAlaAsnValHisGlnGluAsnGluGluMetGluGlnProLeuGlnAs 17
|||||
312  ATGGCAATATTACCAGAAACGAGAGATGGAGCCTATGCAGAA 361

17  nGlyGluGluAspArgProValGlyGlyGlyGlyGlyHisGlnProAlaG 34
|||||
362  TGCAGAGGAGACCGCCCTTTGGGAGGAGGTGAAGCCACCAGCCTCCAG 411

34  lyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArgArg 50
|||||
412  GAAAT.....CGACGG 422

51  GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67
|||||
423  GGACAGGCTGCGCGACTTGGCCCTTAATTTTCGATGGCCCATACCAATAG 472

67  gGlnMetAsnAspGlyLeuGlyArgGlyAspGlyAspMetGluMetPheM 84
|||||
473  GCAGATCAATGATGGATGGGTGGAGATGGAGATGATGTAATATTCATCA 522

84  etGluGluMetArgGluLeuArgArgLysLeuArgGluLeuGlnLeuArg 100
|||||
523  TGGAGGAGATGAGAGAAATCAGAAGAACTTGGGAGCTGCAGTTCCAG 572

101  AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH 117
|||||
573  AATTGTCTGCGTATCCTTATGGGGAGCTCTCTAATCACCATGACCATCA 622

117  sAspGluPheCysLeuMetPro 124
|||||
623  TGATGAATTTTGCCCTTATGCCT 644

```

seq\_name: gb\_pr:AF187064

```

seq_documentation_block:
LOCUS AF187064 891 bp mRNA 11-JUN-2000
DEFINITION Homo sapiens p75NTR-associated cell death executor (NADE) mRNA,
complete cds.
ACCESSION AF187064
VERSION AF187064.1 GI:8452893
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
Hanaoka,T., Li,Y., Irie,S. and Sato,T.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES
    source
        1..891

```



```
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .891
/gene="NADE"
312. .647
/gene="NADE"
/note="Involved in the common neurotrophin receptor
p75NTR-mediated signal transduction"
/codon_start=1
/product="p75NTR-associated cell death executor"
/protein_id="AAF75129.1"
/db_xref="GI:8452894"
/translation="WANIHQENEEMEQPNQGEEDRPLGGEGHQPAGNRRQARRLA
PNFRWAIPLNRQINDMGDDMEIFMEEMREIRKLRELQLRNLRLMGLSNHHD
HHDEFCLAP"
BASE COUNT 251 a 182 c 224 g 234 t
ORIGIN
1 MetAlaSnValHisGlnGluAsnGluMetGluProLeuGlnAs 17
312 ATGGCAATATTACACAGGAAAGAGAGATGGAGCGCTATGCAGAA 361
17 nGlyGluAspArgProValGlyGlyGlyGlyHisGlnProAlaG 34
362 TGGAGAGAGACCGCCCTTGGGAGGAGGTGAAGGCCACCGCTGCAG 411
34 lyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnArg 50
412 GAAT.....CGACGG 422
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67
423 GGACAGGCTCGCCGACTTGCCTTAATTTTCGATGGGCCATACCAATAG 472
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
473 GCATCAATGATGGGATGGGTGGAGATGGAGATGATGGAATATCA 522
84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100
523 TGGAGGAGATGAGAGAAATCAGAGAAAACCTTAGGGAGCTGCAGTTCAGG 572
101 AsnCysLeuArgIleLeuMetGlyLeuSerAsnHisAspHisH 117
573 AATTTCTGCTATCCTTATGGGGAGCTCTCTAATCACCATTGACCATCA 622
117 sAspGluPheCysLeuMetPro 124
623 TGATGAATTTGCCTTATGCCT 644
seq_name: gb_sts:G72708
seq_documentation_block: 421 bp DNA STS 08-AUG-2001
LOCUS G72708
DEFINITION MARC 4953-4954:991939031:1 SCF - porcine spleen Sus scrofa STS
genomic, sequence tagged site.
ACCESSION G72708
VERSION G72708.1 GI:15146738
KEYWORDS STS.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE 1 (bases 1 to 421)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L. and
Keeler,J.W.
TITLE Single nucleotide polymorphism (SNP) discovery in expressed porcine
genes
JOURNAL Unpublished (2001)
COMMENT
Contact: Freking BA
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4278
Fax: 402 762 4173
Email: freking@email.marc.usda.gov
Primer A: GCAATGGTCAAACTCTACT
Primer B: CCAGCAGCAATAGACG
STS size: 500
PCR Profile:
Hotstart: 95 degrees for 15 minutes
Denature: 95 degrees for 30 seconds
Anneal: 56 degrees
Extension: 68 degrees for 2 minutes
Cycles: 32 to 45
Protocol:
Template: 50-200 ng genomic DNA
Primer: each 20 pmoles
dNTPs: each 88 uM
Taq Polymerase: 0.25 units (Qiagen HotStar)
Buffer: Commercially supplied Qiagen HotStar buffer
The STS is derived from PCR amplicons generated from genomic DNA,
sequenced from each end using the amplification primers. The
sequence does not necessarily represent the entire amplicon.
Sequence derived from PolyPhred was trimmed from each end of each
unique contig until five consecutive bases exceeded a quality score
threshold of 20, and the next 10 bases averaged a quality score of
20 or greater. Amplicon size was estimated by agarose gel
electrophoresis.
FEATURES
Location/Qualifiers
1. .421
/organism="Sus scrofa"
/strain="white composite, duroc, meishan, minzhu,
fengjing, crossbreds"
/db_xref="taxon:9823"
/sex="male and female"
/clone_lib="SCF - porcine spleen"
/dev_stage="adult"
/note="Organ: spleen"
<1. >421
BASE COUNT 130 a 101 c 86 g 102 t 2 others
ORIGIN
alignment_scores:
Quality: 411.00 Length: 77
Ratio: 5.338 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.104
alignment_block:
US-09-327-750D-12 x G72708/rev ..
Align seg 1/1 to reverse of: G72708 from: 1 to: 421
48 HisArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIle 64
:::|||||
421 AATAGACGGGCACAAAGCTGCGCGACTTGCCTTAATTTCCGATGGCCAT 372
64 eProAsnArgGlnMetAsnAspGlyLeuGlyAspGlyAspMetG 81
|||||
371 ACCCAATAGGCAGATCAATGATGGGTGGATGGAGATGGAGATGATGG 322
```

81 luMetPheMetGluGluMetArgGluIleLeuArgLysLeuArgGluLeu 97  
|||||  
321 AAATGTTTCATGAGAGATGAGAAATCAGAGAAACTTAGGAGCTG 272  
98 GlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisH 114  
|||||  
271 CAGTTGAGGAATGTCGCTATCCTTATGGGGAGCTCTCTAATCACCA 222  
114 sasphishisaspGluPheCysLeuMetPro 124  
|||||  
221 TGACCATCATGATGAATTTGCCTATGCCT 191  
seq\_name: gb\_pr:HSV351F8  
seq\_documentation\_block:  
LOCUS HSV351F8 45678 bp DNA PRI 23-NOV-1999  
DEFINITION Human DNA sequence from cosmid V351F8, between markers DXS366 and  
DXS87 on chromosome X contains ESTs.  
ACCESSION 270719  
VERSION 270719.1 GI:1261915  
KEYWORDS X.  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 45678)  
AUTHORS Whiteley,M.  
TITLE Direct Submission  
JOURNAL Submitted (09-APR-1995) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1RQ, UK. E-mail enquiries: humquerry@sanger.ac.uk  
COMMENT IMPORTANT: This sequence is the entire insert of clone V351F8. The  
true left end of clone V351F8 is at 1 in this sequence. The true  
right end of clone V351F8 is at 45678.  
V351F8 is from the human chromosome X-specific cosmid library.  
FEATURES  
source  
1..45678  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="X"  
/clone="GHC-351F8"  
/clone\_lib="SCCV"  
1..70  
repeat\_region  
/partial  
/note="Alu repeat: matches 80. .1 of consensus"  
642..930  
repeat\_region  
/note="Alu repeat: matches 1. .308 of consensus"  
1406..5179  
repeat\_region  
/note="L1 element fragment"  
5243..5276  
repeat\_region  
/note="17 copies of 2 mer 82 % conserved"  
5279..5567  
/partial  
/note="Alu repeat: matches 304. .1 of consensus"  
9975..10369  
repeat\_region  
/note="MSTA element fragment"  
10399..12017  
repeat\_region  
/note="MSTAR element fragment"  
11318..11978  
repeat\_region  
/note="THE1B element fragment"  
11715..11846  
repeat\_region  
/note="MLT1R element fragment"  
12074..12170  
repeat\_region  
/note="MSTA element fragment"  
12227..12486  
repeat\_region  
/note="MSTA element fragment"  
12437..12502  
repeat\_region  
/note="MSTC element fragment"  
12756..13059  
/partial  
/note="Alu repeat: matches 308. .1 of consensus"  
14478..14650  
repeat\_region  
/note="THE1B element fragment"

repeat\_region  
14491..14536  
/note="MSTA element fragment"  
14715..14831  
repeat\_region  
/note="THE1B element fragment"  
15099..15533  
repeat\_region  
/note="L1 element fragment"  
15534..15626  
repeat\_region  
/note="31 copies of 3 mer 85 % conserved"  
15566..15619  
repeat\_region  
/note="3 copies of 18 mer 98 % conserved"  
16352..16436  
repeat\_region  
/note="L1 element fragment"  
16728..16763  
repeat\_region  
/note="9 copies of 4 mer 94 % conserved"  
16764..17054  
/partial  
repeat\_region  
/note="Alu repeat: matches 308. .1 of consensus"  
19420..19449  
repeat\_region  
/note="15 copies of 2 mer 87 % conserved"  
complement(19537..20013)  
misc\_feature  
/note="match: 3' EST N51315 clone 283089"  
20071..20114  
repeat\_region  
/note="22 copies of 2 mer 98 % conserved"  
20073..20112  
repeat\_region  
/note="10 copies of 4 mer 100 % conserved"  
20073..20126  
repeat\_region  
/note="3 copies of 18 mer 87 % conserved"  
23088..23300  
repeat\_region  
/note="L1 element fragment"  
23311..23490  
repeat\_region  
/note="MLT1A element fragment"  
23318..23489  
repeat\_region  
/note="MLT1B element fragment"  
23362..23491  
repeat\_region  
/note="MLT1C element fragment"  
23507..24407  
repeat\_region  
/note="L1 element fragment"  
24376..24462  
repeat\_region  
/note="MSTC element fragment"  
24379..24491  
repeat\_region  
/note="MSTA element fragment"  
24688..24738  
repeat\_region  
/note="MSTA element fragment"  
24810..25016  
repeat\_region  
/note="L1 element fragment"  
25254..25421  
repeat\_region  
/note="L1 element fragment"  
25439..25696  
repeat\_region  
/note="MER25 element fragment"  
26284..26343  
repeat\_region  
/note="MLT1D element fragment"  
26428..26491  
repeat\_region  
/note="MLT1D element fragment"  
26465..26518  
repeat\_region  
/note="MLT1E element fragment"  
27446..27667  
repeat\_region  
/note="L1 element fragment"  
28954..29247  
/partial  
repeat\_region  
/note="Alu repeat: matches 308. .1 of consensus"  
30160..30325  
repeat\_region  
/note="2 copies of 83 mer 98 % conserved"  
32588..32875  
misc\_feature  
/note="match: 5' EST H68599 clone 239077"  
complement(32825..33230)  
misc\_feature  
/note="match: 3' EST H68239 clone 289077"  
35551..35802  
repeat\_region  
/partial  
/note="Alu repeat: matches 1. .260 of consensus"  
35804..35839  
repeat\_region  
/note="18 copies of 2 mer 83 % conserved"  
38625..38705  
repeat\_region  
/note="MLT1R element fragment"



—  
—  
—  
..  
..  
—  
—  
—  
—  
—  
—





```

repeat_region 15994..16040
repeat_region 16116..16184
repeat_region 16897..17005
repeat_region 17618..17700
repeat_region 17748..17912
repeat_region 18657..18871
repeat_region 18927..19230
repeat_region 19492..19559
repeat_region 19560..19821
repeat_region 19822..19959
repeat_region 19987..20286
repeat_region 20316..20512
repeat_region 20513..20741
repeat_region 20769..21074
repeat_region 21085..21207
repeat_region 21212..21391
repeat_region 21392..21471
repeat_region 22146..22441
repeat_region 22526..22610
repeat_region 22940..23204
repeat_region 23205..23514
repeat_region 23919..24194
repeat_region 24496..24590
repeat_region 24871..24916
repeat_region 25378..25555
repeat_region 25673..25910
repeat_region 26722..26798
repeat_region 27040..27107
repeat_region 27108..27374
repeat_region 27375..27471
repeat_region 27651..27945
repeat_region 28235..28876
repeat_region 28975..29097
repeat_region 29173..29196
repeat_region 29832..30260
misc_feature 30392..30597
repeat_region 30845..31010

```

```

/note="MER91B repeat: matches 2..162 of consensus"
31056..31094
/note="MIR repeat: matches 218..256 of consensus"
31728..31791
/note="16 copies 4 mer t9tg 76 conserved"
31858..31952
/note="MIR repeat: matches 48..153 of consensus"
complement(33437..33589)
/note="match: STS: Em:H5731"
33543..33633
/note="MIR repeat: matches 84..165 of consensus"
33785..33966
/note="MER91A repeat: matches 1..185 of consensus"

```

```

alignment_scores:
  Quality: 240.50      Length: 67
  Ratio: 4.076        Gaps: 2
  Percent Similarity: 88.060  Percent Identity: 77.612

```

```

alignment_block:
US-09-327-750D-12 x HS714B7

```

```

Align seg 1/1 to: HS714B7 from: 1 to: 98274

```

```

56 LeuAlaProAsnPheArgTrpAlaIleProAsnArgGlnMetAsnAspG1 72
|||||
62887 TTGGCCCCCAATTTCAATGGCCCATACCAATAGGAGTCATGATGG 62936

72 yLeuGlyGlyAspGlyAspMetGluMetPheMetGluGluMetArgG 89
|||||
62937 GATCAATAGAGTGGAGATGATATGGAATGTTTCATGGAGAGATGAG 62986

89 luIleArgArg.LysLeuArgGluLeuGluArgAsnCysLeuArgI1 105
|||||
62987 GAATCAGGAGAGAAAT...AAGGAGCTACAATGAGGAATTTGCTGTGTAT 63033

105 eLeuMetGlyGluLeuSerAsnHisAspHisAspGluPheCys 121
|||||
63034 CCTATGGGAAG.CTGTGTAATCCCCATGACCATCATGATGAATTTGCC 63081

```

```

seq_name: gb_sts:G35294

```

```

seq_documentation_block:
LOCUS G35294 477 bp DNA STS
DEFINITION human STS SHGC-37409, sequence tagged site.
ACCESSION G35294
VERSION G35294.1 GI:2459462
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS Myers, R.M.
JOURNAL Unpublished (1997)
COMMENT

```

```

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu

```

```

Primer A: AACATCTTTCATGAAAGTTGATG
Primer B: CTTTGTGGCATCTTCTGCAA
STS size: 106
PCR Profile:

```

```

Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 15 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds

```

REFERENCE  
AUTHORS  
I (bases I  
Peng, Y., Qian

Align seq 1/1 t

seq documentation block:



```

452 CTGAGGAAAAAGCAGTTGAGTCATAGTCTCGCGGCGAGTCAGCAGCTGAC... 499

110 userAsnHisHisAspHisHisAspGluPheCysLeuMetPro 124
      |||
500 .CCCCCTCACCATGACCATCATGATGAGTTTGCCTTATGCGCCC 511

```

**THIS PAGE BLANK (USPTO)**

OM of: US-09-327-750D-12 to: EST:\* out\_format : pfs

Date: Mar 11, 2002 2:16 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=Cgnt2\_1/USPTO.spool/US09327750/runat\_11032002\_101153\_20308/app\_query.fasta\_1.1472  
-DB=EST\_QPWT-fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPECL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODEL=LOCAL -OUTFWT=pfs -NORM=ext -HEAPSIZ=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09327750\_ECGN1\_1.5654  
-NCPU=6 -ICPU=3 -IONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-12

Query length: 124

Database: EST:

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 4085.940000

score\_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_est2:W46041	+	688.00	1165.70	9.1e-56	492	!
gb_est1:AI006575	+	688.00	1164.93	1.0e-55	538	!
gb_est1:BE334866	+	688.00	1164.87	1.0e-55	542	!
gb_est1:AA215070	+	688.00	1164.63	1.0e-55	557	!
gb_est1:BE334877	+	688.00	1164.21	1.1e-55	585	!
gb_est2:W64711	+	688.00	1164.13	1.1e-55	590	!
gb_est1:BG088461	+	688.00	1164.00	1.1e-55	599	!
gb_est1:AA223275	+	688.00	1163.82	1.2e-55	612	!
gb_est2:W61757	+	688.00	1163.76	1.2e-55	616	!
gb_est1:AW47468	+	688.00	1163.60	1.2e-55	628	!
gb_hic:AK004531	+	688.00	1161.41	1.6e-55	809	!
gb_est1:AA253897	+	687.00	1164.44	1.1e-55	468	!
gb_hic:AK010500	+	683.00	1152.56	4.9e-55	845	!
gb_est1:AI152323	+	681.00	1155.00	3.6e-55	430	!
gb_est1:AW08751	+	679.00	1150.26	6.6e-55	503	!
gb_est2:BF681983	+	675.00	1149.08	7.7e-55	577	!
gb_est1:AA268306	+	675.00	1142.30	1.8e-54	577	!
gb_est1:AI118980	+	673.00	1141.38	2.1e-54	642	!
gb_est2:BF012441	+	673.00	1141.45	2.0e-54	430	!
gb_est1:AA636864	+	673.00	1139.99	2.5e-54	509	!
gb_est2:BF608455	+	673.00	1136.23	4.0e-54	787	!
gb_est1:AA475682	+	667.00	1131.70	7.2e-54	410	!
gb_est1:AA044548	+	661.00	1120.01	3.2e-53	489	!
gb_est1:AA466788	+	661.00	1119.15	3.6e-53	540	!
gb_est2:BI407164	+	661.00	1115.58	5.7e-53	817	!
gb_est2:W98871	+	659.00	1115.83	5.5e-53	536	!
gb_est1:AA667520	+	658.00	1115.09	6.0e-53	480	!
gb_hic:AK003294	+	657.00	1108.71	1.4e-52	826	!
gb_est1:AA004191	+	652.00	1107.26	1.6e-52	542	!
gb_est1:AA050176	+	654.00	1103.86	2.5e-52	543	!
gb_est1:AA606578	+	650.00	1100.63	3.8e-52	533	!
gb_est1:AW476461	+	645.00	1091.93	1.2e-51	547	!
gb_est1:AA049613	+	642.00	1087.52	2.1e-51	506	!
gb_est1:BE198421	+	640.00	1084.98	2.9e-51	459	!
gb_est1:BE198421	-	640.00	1082.62	3.9e-51	603	!
gb_est2:BI082462	+	639.00	1077.93	7.1e-51	853	!
gb_est1:AA214902	+	637.00	1078.60	6.5e-51	533	!
gb_est1:AA259786	+	634.00	1074.32	1.1e-50	486	!
gb_est1:AA110311	+	634.00	1074.07	1.2e-50	500	!
gb_est1:AA068855	+	631.00	1068.62	2.3e-50	522	!

gb\_est2:BI281721 - 628.00 1061.48 5.8e-50 662 ! BI281721 UI-R-CT0s-cav-f-10-  
gb\_est2:BF168928 + 621.50 1048.73 3.0e-49 810 ! BF168928 60177532F1 NCI\_CGA  
gb\_est1:AA214909 + 618.00 1046.43 4.0e-49 532 ! AA214909 mu76f05.rl Stratagene  
gb\_est1:AI183075 + 615.00 1042.69 6.5e-49 455 ! AI183075 ub93e07.rl Soares\_m  
gb\_est1:AI157489 + 611.00 1037.80 1.2e-48 366 ! AI157489 ue56e05.rl Soares\_m

seq\_name: gb\_est2:W46041

seq\_documentation\_block:

LOCUS W46041 mRNA 23-MAY-1996

DEFINITION mc82h02.rl Soares mouse embryo NBME13.5 14.5 Mus musculus CDNA

clone IMAGE:355059 5' similar to gb:M38188 OVARIAN GRANULOSA CELL

13.0 KD PROTEIN HGR74 (HUMAN); mRNA sequence.

ACCESSION W46041

VERSION W46041.1 GI:1330757

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 492)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LML; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:226859

Seq primer: ETPrimer

High quality sequence stop: 489.

Location/Qualifiers

1..492

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:355059"

/clone\_lib="Soares mouse embryo NBME13.5 14.5"

/sex="unknown"

/tissue\_type="embryo"

/dev\_stage="13.5-14.5dpc total fetus"

/lab\_host="DH10B"

/note="Vector: pTT73D-Pac (Pharmacia) with a modified

polylinker; Site1: Not I; Site2: Eco RI; 1st strand CDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCGAATGGCGCGGAAATTTTTTTTTTTTTTTTTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos [total RNA provided by Minoru Ko, Wayne

State Univ., from 2 1; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pTT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M.Fatima Bonaldo."

BASE COUNT 133 a 133 c 125 g 101 t

ORIGIN

alignment\_scores:

Quality: 688.00 Length: 124

Ratio: 5.548 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.387

alignment\_block:

US-09-327-750D-12 x W46041 ..

Align seg 1/1 to: W46041 from: 1 to: 492

```

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
13 ATGGCCAAATGTCACCAAGAAACGAGAGATGGAGCAGCCCTGCAGAA 62
17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlyGlnProAlaG 34
63 TGGACAGGAGACCGCCCTGTGGAGAGGTGAGGGCCACCAAGCTGCTG 112
34 lyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAspArg 50
113 CAACAACAACAACAACAACAACAACCAATACCAACCAACCAACCAAGA 162
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsn 67
163 GCCCAGGCTCGCGGACTTGCCTTAACCTTCCGATGGGCCATTCACCA 212
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
213 GCAGATGATGACGGTGGTGGAGATGGAGATGATATGGAAATGTTCA 262
84 etGluGluMetArgGluIleArgLysLeuArgGluGlnLeuArg 100
263 TGGAGGAGATGAGAGATCGGAGAAAGCTTAGGAGCTACAGCTGAGA 312
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisAspHisH 117
313 AATTGCTACGCATCTTTATGGGGAGCTGTCTAACCAACCAACCAACCA 362
117 sAspGluPheCysLeuMetPro 124
363 TGATGAATCTGCCTTATGCCT 384

```

seq\_name: gb\_estl:A1006575

```

seq_documentation_block: 538 bp mRNA EST 12-JUN-1998
LOCUS A1006575
DEFINITION uel5e06.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1480450 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION A1006575
VERSION A1006575
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 538)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le.M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilsson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:928806
Seq primer: primer name ambiguous
High quality sequence stop: 433.
Location/Qualifiers
1..538
/organism="Mus musculus"

```

```

/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1480450"
/clone_lib="Sugano mouse embryo mewa"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"
/note="vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGGCTTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCTGCTCTAAAGCTGCG and 3' end primer
CGACCTGCAGCTCGACCA."
BASE COUNT 154 a 140 c 162 g 82 t
ORIGIN

```

```

alignment_scores:
Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387

```

alignment\_block:

US-09-327-750D-12 x A1006575 ..

Align seg 1/1 to: A1006575 from: 1 to: 538

```

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
162 ATGGCCAAATGTCACCAAGAAACGAGAGATGGAGCAGCCCTGCAGAA 211
17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlyGlnProAlaG 34
212 TGGACAGGAGACCGCCCTGTGGAGAGGTGAGGGCCACCAAGCTGCTG 261
34 lyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAspHisArg 50
262 CAACAACAACAACAACAACAACAACCAACCAACCAACCAACCAAGA 311
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsn 67
312 GCCCAGGCTCGCGGACTTGCCTTAACCTTCCGATGGGCCATTCACCA 361
67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84
362 GCAGATCAATGACGGTGGTGGAGATGGAGATGATATGGAAATGTTCA 411
84 etGluGluMetArgGluIleArgLysLeuArgGluGlnLeuArg 100
412 TGGAGGAGATGAGAGATCCGGAGAAAGCTTAGGGAGCTACAGCTGAGA 461
101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisAspHisH 117
462 AATTGCTACGCATCTTTATGGGGAGCTGTCTAACCAACCAACCAACCA 511
117 sAspGluPheCysLeuMetPro 124
512 TGATGAATCTGCCTTATGCCT 533

```

seq\_name: gb\_estl:BE334866

seq\_documentation\_block:

```

LOCUS BE334866
DEFINITION us90a11.y1 Soares mammary_gland_NWLMG Mus musculus cDNA clone
IMAGE:3325628 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION BE334866

```

```
VERSION BE334866.1 GI:9208642
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1069792
Seq primer: -40RP from Gibco
High quality sequence stop: 465.
Location/Qualifiers
1..542
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3325628"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 156 a 141 c 164 g 81 t
ORIGIN

alignment_scores
Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387

alignment_block
US-09-327-750D-12 x BE334866 ..
Align seg 1/1 to: BE334866 from: 1 to: 542

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
166 ATGCCCAATCTCCACCAGAAAGAGAGATGGAGCAGCCCTGCAGAA 215
17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlyHisGlnProAlaG 34
216 TGGACAGAGAGACGGCCTCTGGGAGAGGTGAGGGCCAGCAGCTGCTG 265
34 lYAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArg 50
266 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 315
51 GlyGlnAlaArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67
316 GGCAGAGCTCCGACACTGCCCTTAACCTCCGATGGCCATCCCAACAG 365
67 gGlnMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 84
366 GCAGATGAATGACGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGT 415
84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100
416 TGGAGGAGATGAGAGATCGGAGAAAGCTTAGGGAGCTACAGCTGAGA 465

seq_name: gb_estl:AA215070

seq_documentation_block:
LOCUS AA215070 557 bp mRNA EST 03-FEB-1997
DEFINITION mu76f04.r1 Stratagene mouse embryonic carcinoma (#937317) Mus
musculus cDNA clone IMAGE:651487 5' similar to gb:M38188 OVARIAN
GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION AA215070
VERSION AA215070.1 GI:1814831
KEYWORDS EST:
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 557)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:397335
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 447.
Location/Qualifiers
1..557
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:651487"
/clone_lib="Stratagene mouse embryonic carcinoma (#937317)
"
/tissue_type="carcinoma"
/dev_stage="embryonic"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. p19 cell
line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
BASE COUNT 159 a 145 c 165 g 88 t
ORIGIN

alignment_scores
Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.387

alignment_block
US-09-327-750D-12 x AA215070 ..
Align seg 1/1 to: AA215070 from: 1 to: 557

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
|||||
```

141 ATGGCAATGTCACCAAGAAACGAGAGATGGAGCACCCTCGAGAA 190  
 17 nGlyGluAspArgProValGlyGlyGluGlyHisGlnProAlaG 34  
 191 TGGACAGGAAGACCGCCCTGTGGAGGAGGTGAGGCCACCACTGCTG 240  
 34 lyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArgArg 50  
 241 CAAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 290  
 51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAlaIleProAsnAr 67  
 291 GCCAGGCTCGCGGCTTCCCTTAACTTCCGATGGCCATTCCCAACAG 340  
 67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84  
 341 GCAGATGAATGACGGGTGGGTGGAGATGGAGATGATGAAATGTTCA 390  
 84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100  
 391 TGGAGGAGATGAGAGAGATCCGAGAAAGCTTAGGGAGCTACAGCTGAGA 440  
 101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH 117  
 441 AATTGCTACGATCCTTATGGGGAGCTGTCTAACCAACCAAGATCACCA 490  
 117 sAspGluPheCysLeuMetPro 124  
 491 TGATGAATTCGCCCTTATGCCT 512

seq\_name: gb\_est1:BE334877

seq\_documentation\_block: 14-JUL-2000  
 LOCUS BE334877 585 bp mRNA EST  
 DEFINITION us90b10.y1 Soares mammary gland\_NMLMG Mus musculus cDNA clone IMAGE:3325627 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.  
 ACCESSION BE334877 GI:9208653  
 VERSION BE334877.1  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 585)  
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
 This clone is available royalty-free through LILNL; contact the IMAGE Consortium ([info@image.lilnl.gov](mailto:info@image.lilnl.gov)) for further information.  
 MGI:1069791  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 465.

## FEATURES

## Source

1..585  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_image="3325627"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 162 a 157 c 172 g 93 t 1 others  
 ORIGIN

## alignment\_scores:

Quality: 688.00 Length: 124  
 Ratio: 5.548 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.387

## alignment\_block:

US-09-327-750D-12 x BE334877 ..

Align seg 1/1 to: BE334877 From: 1 to: 585

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17  
 166 ATGGCCAATGTCACCAAGAAACGAGAGATGGAGCACCCTCGAGAA 215  
 17 nGlyGluGluAspArgProValGlyGlyGlyGluGlyHisGlnProAlaG 34  
 216 TGGACAGGAAGACCGCCCTGTGGAGGAGGTGAGGCCACCACTGCTG 265  
 34 lyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArgArg 50  
 266 CAAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 315  
 51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAlaIleProAsnAr 67  
 316 GCCAGGCTCGCGGCTTCCCTTAACTTCCGATGGCCATTCCCAACAG 365  
 67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84  
 366 GCAGATGAATGACGGGTGGGTGGAGATGGAGATGATGAAATGTTCA 415  
 84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100  
 416 TGGAGGAGATGAGAGAGATCCGAGAAAGCTTAGGGAGCTACAGCTGAGA 465  
 101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH 117  
 466 AATTGCTACGATCCTTATGGGGAGCTGTCTAACCAACCAAGATCACCA 515  
 117 sAspGluPheCysLeuMetPro 124  
 516 TGATGAATTCGCCCTTATGCCT 537

seq\_name: gb\_est2:w64711

seq\_documentation\_block:

LOCUS W64711 590 bp mRNA EST 10-JUN-1996  
 DEFINITION md71q01.r1 Soares mouse embryo NbME13 5 14.5 Mus musculus cDNA clone IMAGE:373872 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

## ACCESSION

W64711

## VERSION

W64711.1 GI:1372353

## KEYWORDS

EST.

## SOURCE

house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 590)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

## TITLE

The WashU-HIMI Mouse EST Project

## JOURNAL

Unpublished (1996)

## COMMENT

Contact: Marra M/Mouse EST Project  
 WashU-HIMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800

Fax: 314 286 1810  
 Email: mouse@wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:235304

Seq primer: EFPrimer  
 High quality sequence stop: 348.

#### FEATURES

Source  
 Location/Qualifiers  
 1..590  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:373872"  
 /clone\_lib="Soares mouse embryo NbME13.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"  
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5'  
 TCTTACCAATCTGAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTT  
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
 State Univ., from 2 ]; double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 pT73 vector. Library went through one round of  
 normalization, and was constructed by Bento Soares and  
 M.Fatima Bonaldo."

BASE COUNT 165 a 155 c 166 g 104 t  
 ORIGIN

#### alignment\_scores:

Quality: 688.00 Length: 124  
 Ratio: 5.548 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.387

#### alignment\_block:

US-09-327-750D-12 x W64711

Align seg 1/1 to: W64711 from: 1 to: 590

1 MetAlaAsnValHisGlnGluAsnGluMetGluProLeuGlnAs 17  
 134 ATGCCCAATGTCCACGAGGAAACGAGAGATGGAGCCCTGCAGAA 183  
 17 nGlyGluGluAspArgProValGlyGlyGlyHisGlnProAlaG 34  
 184 TGGACAGGAAGACCGCCTGTGGGAGGAGTGAGGGCCACAGCCTGCTG 233  
 34 lyAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisHisArgArg 50  
 234 CAACACAAACAACAACAACCAACCAACCAACCAACCAACCAACCA 283  
 51 GlyGluAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67  
 284 GGCCAGGCTCGCCGACTTGGCCCTAACTCCGATGGGCCATCCCAACAG 333  
 67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84  
 334 GCAGATGAATGACGGTGTGGTGAGATGGAGATGATATGGAATGTCA 383  
 84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100  
 384 TGGAGGAGATGAGAGAGATCCGAGAAAGCTTAGGGAGCTACAGCTCAGA 433  
 101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH 117  
 434 AATGTCTACGCATCCCTTATGGGGAGCTGTCTAAACCACCACGATCACCA 483  
 117 sAspGluPheCysLeuMetPro 124

|||||  
 484 TGATGAATTCGTGCTTATGCT 505

seq\_name: gb\_est2:BG088461

#### seq\_documentation\_block:

LOCUS BG088461 599 bp mRNA EST 26-JAN-2001  
 DEFINITION H3153D07-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
 H3153D07 5', mRNA sequence.  
 ACCESSION BG088461  
 VERSION BG088461.1 GI:12571025  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 599)  
 AUTHORS Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka  
 ,T.S., Carter,M.G. and Ko,M.S.H.  
 TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_ESTs: H3153D07-3  
 Contact: George J. Kargul  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgsun.grc.nia.nih.gov  
 This clone set has been freely distributed to the community. Please  
 visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.  
 Plate: H3153 row: D column: 07  
 Seq primer: -21M13 Reverse  
 High quality sequence stop: 599  
 POLYA-No.

#### FEATURES

##### Location/Qualifiers

1..599  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="H3153D07"  
 /clone\_lib="NIA Mouse 15K cDNA Clone Set"  
 /sex="Clones arrayed from a variety of cDNA libraries"  
 /dev\_stage="Clones arrayed from a variety of cDNA  
 libraries"  
 /lab\_host="DH10B"  
 /note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; This  
 clone is among a rearranged set of 15,247 clones from 11  
 embryo cDNA libraries (including preimplantation stage  
 embryos from unfertilized egg to blastocyst, embryonic  
 part of E7.5 embryos, extraembryonic part of E7.5 embryos  
 , and E12.5 female mesonephros/gonad) and one newborn  
 ovary cDNA library. Average insert size 1.5 kb. All  
 source libraries are cloned unidirectionally with Oligo(dT  
 )-Not primers. References include: (1) Genome-wide  
 expression profiling of mid-gestation placenta and embryo  
 using a 15,000 mouse developmental cDNA microarray, 2000,  
 Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)  
 Large-scale cDNA analysis reveals phased gene expression  
 patterns during preimplantation mouse development, 2000,  
 Development, 127: 1737-1749; (3) Genome-wide mapping of  
 unselected transcripts from extraembryonic tissue of  
 7.5-day mouse embryos reveals enrichment in the t-complex  
 and under-representation on the X chromosome, 1998, Hum  
 Mol Genet 7: 1967-1978."

BASE COUNT 163 a 163 c 176 g 97 t  
 ORIGIN

#### alignment\_scores:

Quality: 688.00 Length: 124  
 Ratio: 5.548 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.387

#### alignment\_block:

```
US-09-327-750D-12 x BG088461
Align seg 1/1 to: BG088461 from: 1 to: 599
1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
175 ATGGCCCAATGTCACACGAGAAACGAGAGATGGAGCCCTGCAGAA 224
17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlnProAlaG 34
225 TGGACAGGAAGACCCCTGTGGAGGAGGTGAGGCCACACGCTGCTG 274
34 lYAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHisArg 50
275 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 324
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaLeuProAsnAr 67
325 GGCCAGGCTCGCGGACTTGCCTTAACCTTCGATGGGCCATTCACACAG 374
67 gGlnMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGly 84
375 GCAGATGAATGACGGCTTGGGTGGAGATGGAGATGATGAAATGTTCA 424
84 etGluGluMetArgGluLeuArgArgLysLeuArgGluGlnLeuArg 100
425 TGGAGGAGATGAGAGAGATCCGGAAGAGCTTAGGGAGCTACAGCTGAGA 474
101 AsnCysLeuArgIleLeuMetGlyGlyGlyGlyGlyGlyGlyGlyGly 117
475 AATTGCTACGCATCTTATGGGGAGCTGTCTAACCCACACGATCACCA 524
117 sAspGluPheCysLeuMetPro 124
525 TGATGAATTCGCCTTATGCCT 546
seq_name: gb_est1:AA272375

seq_documentation_block:
LOCUS AA272375 612 bp mRNA EST 26-MAR-1997
DEFINITION vb62407.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
clone IMAGE:761628 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION AA272375
VERSION AA272375.1 GI:1910706
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE
AUTHORS
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL: contact the
IMAGE Consortium (info@image.lnli.gov) for further information.
MGI:462548
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 507.
Location/Qualifiers
1, 612
/organism="Mus musculus"

/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:761628"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACGAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[GTGGATTCCGTACC], digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."
BASE COUNT 168 a 167 c 175 g 102 t
ORIGIN

alignment_scores:
Quality: 688.00 Length: 124
Ratio: 5.548 Gaps: 0
Percent Similarity: 100.000 Percent identity: 98.387
alignment_block:
US-09-327-750D-12 x AA272375
Align seg 1/1 to: AA272375 from: 1 to: 612
1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17
176 ATGGCCAATGTCACACGAGAAACGAGAGATGGAGCCCTGCAGAA 225
17 nGlyGluGluAspArgProValGlyGlyGlyGlyGlnProAlaG 34
226 TGGACAGGAAGACCCCTGTGGAGGAGGTGAGGCCACACGCTGCTG 275
34 lYAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisHisArg 50
276 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 325
51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaLeuProAsnAr 67
326 GGCCAGGCTCGCGGACTTGCCTTAACCTTCGATGGGCCATTCACACAG 375
67 gGlnMetAsnAspGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGly 84
376 GCAGATGAATGACGGCTTGGGTGGAGATGGAGATGATGAAATGTTCA 425
84 etGluGluMetArgGluLeuArgArgLysLeuArgGluGlnLeuArg 100
426 TGGAGGAGATGAGAGAGATCCGGAAGAGCTTAGGGAGCTACAGCTGAGA 475
101 AsnCysLeuArgIleLeuMetGlyGlyGlyGlyGlyGlyGlyGlyGly 117
476 AATTGCTACGCATCTTATGGGGAGCTGTCTAACCCACACGATCACCA 525
117 sAspGluPheCysLeuMetPro 124
526 TGATGAATTCGCCTTATGCCT 547
seq_name: gb_est2:W81757

seq_documentation_block:
LOCUS W81757 616 bp mRNA EST 12-SEP-1996
DEFINITION me95d06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:403307 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION W81757
VERSION W81757.1 GI:1392776
KEYWORDS EST.
SOURCE house mouse.
```



## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 616)

**AUTHORS**  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:247075

Putative full length read

vector to vector length is 617

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 492.

Location/Qualifiers

## FEATURES

source

1..616

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="IMAGE:403307"

/clone\_lib="Soares mouse embryo NbMEL3.5 14.5"

/sex="unknown"

/tissue\_type="embryo"

/dev\_stage="13.5-14.5dpc total fetus"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

## BASE COUNT

ORIGIN

159 a 159 c 159 g 139 t

## alignment\_scores:

Quality: 688.00 Length: 124

Ratio: 5.548 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.387

## alignment\_block:

US-09-327-750d-12 x W81757

Align seg 1/1 to: W81757 from: 1 to: 616

1 MetAlaAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17

|||||

7 ATGCCCAATGCCACCAAGAAACGAGAGATGGAGCCCTGCCAGAA 56

|||||

17 nGlyGluAspArgProValGlyGlyGlyHisGlnProAlaG 34

|||||

57 TGGACAGGAACGCCCTCTGGGAGGAGGTGAGGCCACCGCTGCTG 106

|||||

34 lyAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisArg 50

|||||

107 CAACAACAACAACAACAACAACAACAACAACAACAACAACA 156

51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67

|||||

157 GCCCAGGCTCCCGACATTGCCCTAACTTCCGATGGCCATCCCAACAG 206

|||||

67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84

|||||

207 GCAGATGAATCACCGGTTGGGTGGAGATGGAGATGATGAAATGTTCA 256

|||||

84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100

|||||

257 TGGAGGAGATGAGAGATCCGAGAAAGCTTAGGGAGCTACAGCTGAGA 306

|||||

101 AsnCysLeuArgIleLeuMetGlyGlyLeuSerAsnHisHisAspHisH 117

|||||

307 AATTGTCTACCATCCTTATGGGGAGCTGTCTAACCAACCAGGATCACCA 356

|||||

117 sAspGluPheCysLeuMetPro 124

|||||

357 TGATGAATCTGCCTTATGCCT 378

seq\_name: gb\_estl:AW476468

seq\_documentation\_block:

LOCUS AW476468 628 bp mRNA EST 24-FEB-2000

DEFINITION uc76d02.y1 NCI\_CGAP\_Lu33 Mus musculus cDNA clone IMAGE:2937027 5'

similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74

(HUMAN); mRNA sequence.

ACCESSION AW476468

VERSION AW476468.1 GI:7046574

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 628)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps@email.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, at:

www.llnl.gov/bbrp/image/image.htm!

MG: 2937027

Seq primer: -40RP from Gibco

High quality sequence stop: 456.

Location/Qualifiers

## FEATURES

source

1..628

/organism="Mus musculus"

/strain="CZECH II"

/db\_xref="taxon:10090"

/clone\_lib="IMAGE:2937027"

/clone\_lib="NCI\_CGAP\_Lu33"

/tissue\_type="pooled lung tumors"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; Site\_1: NotI; Site\_2: EcoRI; 1st

strand cDNA was prepared from mRNA obtained from pooled

lung tumors with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not

I and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 171 a 162 c 181 g 112 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 688.00 Length: 124  
Ratio: 5.548 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.387  
alignment\_block:  
US-09-327-750d-12 x AW476468 ..

Align seg 1/1 to: AW476468 from: 1 to: 628

1 MetaAAsnValHisGlnGluAsnGluMetGluGlnProLeuGlnAs 17  
|||||  
138 ATGGCCAAATGTCACACGAGAAACGAGAGATGGAGCCCTCGCAGAA 187  
|||||  
17 nGlyGluGluAspArgProValGlyGlyGlyGluGlyHisGlnProAlaG 34  
|||||  
188 TGCACAGGAAGACCCGCTGTGGGAGAGGTGAGGCCACCGCCTGCTG 237  
|||||

34 lYAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHisArg 50  
|||||  
238 CAAC 287  
|||||

51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 67  
|||||  
288 GCCCAGGTCGCGGACTTGCCTTAACCTCCGATGGCCATCCCAACAG 337  
|||||

67 gClnMetAsnAspGlyLeuGlyGlyAspGlyAspMetGluMetPheM 84  
|||||  
338 CCAGATGAATGACGGGTGGGTGGAGATCGAGATGATGGAATGTTCA 387  
|||||

84 eGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 100  
|||||  
398 TGGAGGAGATGAGAGAGATCGGAGAAAGCTTAGGGAGCTACAGCTGAGA 437  
|||||

101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH 117  
|||||  
438 AATTGCTACGCATCTTATGGGGAGCTGTCTAACCCACCGATCACCA 487  
|||||

117 sAspGluPheCysLeuMetPro 124  
|||||  
488 TGATGAATCTGCCTTATGCCT 509  
|||||

seq\_name: gb\_hlc-AK004531

seq\_documentation\_block:  
LOCUS AK004531 809 bp mRNA HTC 05-JUL-2001  
DEFINITION Mus musculus 18 days embryo cDNA, RIKEN full-length enriched library, clone:1190011J23, full insert sequence.  
ACCESSION AK004531  
VERSION AK004531.1 GI:12835760  
KEYWORDS CAP trapper.  
SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library  
clone:1190011J23.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 809)  
AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning  
JOURNAL Methods in enzymology. 303, 19-44 (1999)  
MEDLINE 99279253

PUBMED 10349636

REFERENCE 2 (bases 1 to 809)

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome research. 10 (10), 1617-1630 (2000)

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

20499374  
11042159

3 (bases 1 to 809)

Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome research. 10 (11), 1757-1771 (2000)

20530913

11076861

4 (bases 1 to 809)

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 809)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Horl,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
Direct Submission

TITLE  
JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGAGAGAGCGCCGCACTCGAGTTTCTTTTCTTTT 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence 5' GAGAGAGAGAGGATCCAGAGCTCAATTAATTAATAACCCCTCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

source

Location/Qualifiers

1..809  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGD:MGI:1338016"  
/db\_xref="MGD:MGI:1892692"  
/clone="1190011J23"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="18 days embryo"  
181..624  
/note="putative"  
/codon\_start=1  
/protein\_id="BAB23350.1"  
/db\_xref="GI:12835761"  
/translation="MANVQENDEMEQPLQNGQEDRPVGGGEGHQPAANNNNNNHNNHNNHRRGQARRLAPNFWAIPNRQMDNGLGGDDMMFMEEMRELRRLRLQLRNC LRLMLGELSNHHDDHDEFCLMPWLSRFPPEIHTVTTPAVALFAFS"

CDS





101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisH1 117  
 478 AATTGTCTACGCATCTTATGGGAGCTGTCTAAACCACCGATCACCA 527  
 117 saspGluPheCysLeuMetPro 124  
 528 TGATGAATTCTGCCTTATGCCT 549

seq\_name: gb\_est1:A1152323

seq\_documentation\_block: 430 bp mRNA EST 30-SEP-1998  
 LOCUS A1152323  
 DEFINITION u079a02.r1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
 IMAGE:1477034 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0  
 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION A1152323

VERSION A1152323.1 GI:3680792

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE  
 AUTHORS  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 430)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL

COMMENT

Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:925390

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 419.

FEATURES

source

1. .430  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1477034"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"

/note="vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT7T3 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Donaldo."

BASE COUNT 135 a 111 c 113 g 71 t

ORIGIN

alignment\_scores:

Quality: 681.00 Length: 124  
 Ratio: 5.537 Gaps: 0  
 Percent Similarity: 99.194 Percent Identity: 97.581

alignment\_block:

US-09-327-750D-12 x A1152323

Align seg 1/1 to: A1152323 from: 1 to: 430

1 MetAlaAsnValHisGlnAsnGluGluMetGluGlnProLeuGlnAs 17  
 54 ATGGCCAATGTCCACAGGAAACGAAGAGATGGAGCAGCCCTGCAGAA 103  
 17 nGlyGluGluAspArgProValGlyGlyGlyGlyHisGlnProAlaG 34  
 104 TGGCAGGAGGAGCCCGCTGTGGATGGAGGTGAGGGCCACCGCCTGCTG 153  
 34 lYAsnAsnAsnAsnAsnAsnHisAsnHisAsnHisAsnHisAsnHis 50  
 154 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 203  
 51 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsn 67  
 204 GCCCAGGCTGCGCGACTTCCCTTAACCTCCGATGGCCCATTCGCCAAC 253  
 67 gGlnMetAsnAspGlyLeuGlyGlyAspGlyAspAspMetGluMetPhe 84  
 254 GCAGATGAATGACGGGTGGGTGGAGATGGAGATGATATGGAATGTTC 303  
 84 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeu 100  
 304 TGGAGGAGATGAGAGAGATCCCGAGAAAGCTTAGGGAGGTACAGCTG 353  
 101 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHis 117  
 354 AATTGTCTACGCATCTTATGGGAGCTGTCTAACCAACCAAGATCACCA 403  
 117 saspGluPheCysLeuMetPro 124  
 404 TGATGAATTCTGCCTTATGCCT 425

seq\_name: gb\_est1:AW908751

seq\_documentation\_block:

LOCUS AW908751

DEFINITION uf57a05.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone

IMAGE:1515440 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0

KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION AW908751

VERSION AW908751.1 GI:8073984

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other ESTs: uf57a05.x1

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:942292

Seq primer: -40RP from Gibco

High quality sequence stop: 468.

FEATURES

source

1. .503

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:1515440"

/clone\_lib="Soares\_mammary\_gland\_NMLMG"

/sex="female (lactating)"

/tissue\_type="mammary gland"

/lab\_host="DH10B"

/note="vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; 1st strand cDNA was prepared from mammary

gland tissue from a lactating female, and was then primed

with a Not I - oligo(dT) primer. Double-stranded cDNA was

ligated to Eco RI adaptors (Pharmacia), digested with Not

I and cloned into the Eco RI and Eco RI sites of the

modified pT7T3 vector. Library is normalized. Library

was constructed by Bento Soares and M. Fatima Donaldo."

I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 132 a 135 c 130 g 105 t 1 others

ORIGIN

alignment\_scores:  
Quality: 679.00 Length: 122  
Ratio: 5.566 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.361

alignment\_block:  
US-09-327-750D-12 x AW908751 ..

Align seg 1/1 to: AW908751 from: 1 to: 503

3 AsnValHisGlnGluAsnGluGluMetGluGlnProLeuGlnAsnGlyG1 19  
|||||  
3 AATGTCCACCAGGAACGAAGAGATGGAGCAGCCCTGCAGATGGACA 52  
19 uGluAspArgProValGlyGlyGluGlyHisGlnProAlaGlyAsnA 36  
:|||||  
53 GGAAGACCGCCCTGTGGGAGGAGGTGAGGGCCACCAGCCTGCTCCAAACA 102  
36 snAsnAsnAsnHisAsnHisAsnHisAsnHisArgArgGlyGln 52  
|||||  
103 ACAACAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAAGGCCAG 152  
53 AlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnArgGlnMe 69  
|||||  
153 GCTCGCGGACTTGCCCTAACTTCGGATGGCCATTCCCAACAGGCAGAT 202  
69 tAsnAspGlyLeuGlyClyAspGlyAspMetGluMetPheMetGluG 86  
|||||  
203 GAATGACGGGTTGGTGGAGATGGAGATGATGAAATGTTTCATGGAGG 252  
86 luMetArgGluIleArgArgGlyLeuArgGluLeuGlnLeuArgAsnCys 102  
|||||  
253 AGATGAGAGAGATCCGGAGAAAGCTTAGGGAGGCTACAGCTGAGAAATGT 302  
103 LeuArgIleLeuMetGlyGluLeuSerAsnHisHisAspHisAspG1 119  
|||||  
303 CTACGCATCCTTATGGGGAGGCTGTCTAACCAACCAACCAACCAACCAATGTA 352  
119 uPheCysLeuMetPro 124  
|||||  
353 ATTCTGCCTTATGCCT 368

OM of: US-09-327-750D-13 to: GenEmbl:\* out\_format : pfs  
Date: Mar 11, 2002 3:33 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=Cgcn2.1/USPTO.spool/US09327750/runat\_11032002\_101154\_20319/app\_query.fasta\_1.1472  
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blomum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTPWT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09327750 -CGN1\_1\_8673  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIME=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-327-750D-13  
Query length: 111  
Database: GenEmbl:\*  
Database sequences: 1472140  
Database length: 34134837  
Search time (sec): 4557.230000

score\_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_pr:BC003190	+	609.00	947.76	1.6e-44	793	BC003190 Homo sapiens, p75NTR-a
gb_pr:HUMOGC	+	609.00	946.94	1.8e-44	891	M38188 Human unknown protein f
gb_pr:AF187066	+	609.00	946.94	1.8e-44	891	AF187064 Homo sapiens p75NTR-as
gb_pr:AF187066	+	567.50	884.08	5.7e-41	700	AF187066 Mus musculus p75NTR-as
gb_ro:AF097440	+	567.50	882.68	6.9e-41	854	AF097440 Mus musculus brain exp
gb_ro:AF187065	+	538.50	841.06	1.4e-38	519	AF187065 Rattus norvegicus p75N
gb_sts:G72708	-	418.00	655.08	3.3e-28	421	G72708 MRC 4953-4954:99193031
gb_pr:HSJ251F8	+	300.00	438.72	3.7e-16	45678	G72708 Human DNA sequence frc
gb_pr:AK000959	+	277.00	428.26	1.4e-15	1229	AK000959 Homo sapiens cDNA FLJ
gb_pr:AX100231	+	277.00	427.53	1.5e-15	1364	AX100231 Sequence 32 from Pat
gb_pr:HS635G19	+	277.00	399.99	5.3e-14	69648	AL0035494 Human DNA sequence f
gb_sts:G35294	-	249.50	392.10	1.5e-13	477	G35294 human SPS SHGC-37409, se
gb_pr:HSJ1847	+	241.50	342.36	8.6e-11	98274	G29755 Human DNA sequence frc
gb_pr:AF183416	+	194.50	304.47	1.1e-08	642	AF183416 Homo sapiens ovarian g
gb_pr:AF220189	+	194.50	302.69	1.4e-08	838	AF220189 Homo sapiens uncharact
gb_pr:HS198P4	+	194.50	276.34	4.1e-07	35714	AL008708 Human DNA sequence f
gb_pr:AF237783	+	187.50	292.12	5.4e-08	791	AF237783 Homo sapiens brain-exp
gb_ro:AF097438	+	175.50	273.51	5.9e-07	785	AF097439 Mus musculus brain exp
gb_pr:AF251053	+	175.00	273.11	6.2e-07	744	AF251053 Homo sapiens X-linked
gb_pr:AX078272	+	175.00	271.79	7.3e-07	898	AX078272 Sequence 76 from Paten
gb_pr:HSV870H8	+	175.00	246.92	1.8e-05	31321	G270233 Human DNA sequence frc
gb_pr:AL133348	-	175.00	245.11	2.2e-05	40584	AL133348 Human DNA sequence f
gb_hlg:HSU80B1	+	175.00	245.03	2.3e-05	41029	AL022169 Homo sapiens chromos
gb_ro:AF097438	+	158.50	246.63	1.8e-05	835	AF097438 Mus musculus brain exp
gb_ro:AF051347	+	158.50	246.61	1.8e-05	838	AF051347 Mus musculus REX-3 mRN
gb_ro:AF097437	+	158.50	239.63	4.5e-05	2269	AF097437 Mus musculus Bex1 pro
gb_pr:HSU105G4	+	156.00	215.00	0.0011	43952	G292846 Human DNA sequence frc
gb_hlg:AL357046	-	140.50	182.08	0.0726	154577	AL357046 Homo sapiens chromo
gb_pr:AC009282	+	140.50	182.07	0.0728	154959	AC009282 Homo sapiens clone
gb_pr:AL35305	+	140.50	180.82	0.0854	185257	AL35305 Human DNA sequence
gb_sts:G24641	-	120.50	191.06	0.0230	504	G24641 human SPS WI-11354, sequ
gb_sts:BC005988	+	105.00	161.09	1.07	1163	BC005988 Homo sapiens, clone 1
gb_pr:HSU105G4	+	105.00	135.67	27.95	43952	G292846 Human DNA sequence frc
gb_ro:BC003254	+	97.00	142.25	12.01	2801	BC003254 Mus musculus, Similar
gb_ba:MXAC000101	+	95.50	138.36	19.78	3622	AC000101 00107, complete sequ
gb_hlg:AC012237	+	95.00	112.55	541.57	129339	AC012237 Homo sapiens clone
gb_hlg:AL592044	+	95.00	108.17	950.81	242092	AL592044 Homo sapiens chromo
gb_pr:AK026349	+	91.50	139.83	16.39	1208	AK026349 Homo sapiens cDNA: FI
gb_pr:HSU177E8	+	91.50	115.18	386.58	40822	G26894 Human DNA sequence frc
gb_pr:HSJ823F3	-	91.50	109.53	797.75	91480	AL079333 Human DNA sequence f

gb\_hlg:LMFLCHR16\_01 - 91.00 107.47 1.0e+03 110000 ! Continuation (2 of 11) of  
gb\_ba:AVIALGEB + 88.50 117.18 299.25 15759 ! L39096 Azotobacter vinelina  
gb\_pl:TIN251864 + 87.50 127.24 82.39 3000 ! AJ251864 Tolypocladium infl  
gb\_ba:AE002479 - 87.50 118.39 256.15 10612 ! AE002479 Neisseria meningi  
gb\_ba:NNAS52491 - 87.50 94.33 5.6e+03 329861 ! AL162756 Neisseria mening  
seq\_name: gb\_pr:BC003190  
seq\_documentation\_block:  
LOCUS BC003190 793 bp mRNA PRI 12-JUL-2001  
DEFINITION Homo sapiens, p75NTR-associated cell death executor; ovarian  
granulosa cell protein (13kD), clone MGC:802 IMAGE:3357965, mRNA,  
complete cds.  
ACCESSION BC003190  
VERSION BC003190.1 GI:13112030  
KEYWORDS MGC.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 793)  
AUTHORS Strausberg, R  
TITLE Direct Submission  
JOURNAL Submitted (13-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
COMMENT Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@cgsc.bc.ca  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAL Plate: 6 Row: k Column: 22  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7657043.  
FEATURES  
Location/Qualifiers  
1..793  
/organism="Homo sapiens"  
/db\_xref="LocusID:27018"  
/db\_xref="taxon:9606"  
/clone="MGC:802 IMAGE:3357965"  
/tissue\_type="Eye, retinoblastoma"  
/clone\_host="DH10B-R"  
/lab\_host="NIH-MGC\_16"  
/note="vector: pOTB7"  
196..531  
/codon\_start=1  
/product="p75NTR-associated cell death executor; ovarian  
granulosa cell protein (13kD)"  
/protein\_id="AAH03190.1"  
/db\_xref="GI:13112031"  
/translation="MANTHOENMEQPMQNGEDRPLGGEGHQPNRRGQARRLA  
PNRWAIPINRQINDMGSGDDGDDMEIFEMREIRLRLQLRNLCLRLMGLSNNHHD  
HHDEFCLMP"  
225 a 170 c 205 g 193 t  
BASE COUNT  
ORIGIN

```
alignment_scores:
  Quality: 609.00      Length: 111
  Ratio: 5.486         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-327-750D-13 x BC003190 ..
  Align seg 1/1 to: BC003190 from: 1 to: 793

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
196 ATGCCAAATATTCACCAAGAAACGAGAGATGGAGCAGCCTATGCAGAA 245

17 nGlyGluGluAspArgProLeuGlyGlyGlyGluGlyHisGlnProAlaG 34
246 TGGAGAGAGAGACCGCCCTTTGGGAGGAGGTGAAGCCACCGCCTGCAG 295

34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
296 GAAATCGACGGGACAGCTCGCGACTTGCCCTAATTTTCGATGGGCC 345

51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
346 ATACCAATAGGCAGATCAATGATGGATGGTGGGAGATGGAGATGATAT 395

67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
396 GGAATATTCATGGAGGAGATGAGAGAAATCAGAAGAAATCTTAGGGAGC 445

84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
446 TGCAGTTGAGGAATGTCGCGTATCTTATGGGGAGCTCTCTAATCAC 495

101 HisAspHisHisAspGluPheCysLeuMetPro 111
496 CATGACCATCATGAATTTTGCCTTATGCCT 528

seq_name: gb_pr:HUOGC

seq_documentation_block:
  LOCUS HUOGC 891 bp mRNA 07-MAR-1995
  DEFINITION Human unknown protein from clone pHR74 mRNA, complete cds.
  ACCESSION M38188 X56942
  VERSION M38188.1 GI:189378
  KEYWORDS
  SOURCE Human ovarian granulosa cell line, cDNA to mRNA.
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 891)
  AUTHORS Rapp,G., Freudenstein,J., Klaudiny,J., Mucha,J., Wempe,F.,
    Zimmer,M. and Scheit,K.H.
  TITLE Characterization of three abundant mRNAs from human ovarian
    granulosa cells
  JOURNAL DNA Cell Biol. 9 (7), 479-485 (1990)
  MEDLINE 91025550
  COMMENT Draft entry and computer-readable sequence for [DNA 9, 479-485
    (1990)] kindly submitted
    by K.H.Scheit, 27-AUG-1990.
  FEATURES
    Location/Qualifiers
      1..891
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="pHR74"
        /cell_type="granulosa"
        /tissue_type="ovary"
        <1..891
        /note="protein of unknown function"
      312..647
        /note="protein of unknown function"
        /codon_start=1
        /protein_id="AAA63232.1"

CDS
  /db_xref="GI:189379"
  /translation="MANIHOENEEMEQPNQCEEDRPLGGEGGHQHPAGNRRRCQARRLA
  PNFRWATPNRQINDMGGGDDMEIFMEEMREIRKRLQLRNLCLRLIMGELSNHHD
  HHDEFCLMP"
  361..534
  /note="protein of unknown function"
  /codon_start=1
  /protein_id="AAA63233.1"
  /db_xref="GI:189380"
  /translation="MERKTALEWEEVKATSLQIDGDLADLPLIFDGPYPPIGRSMWG
  VEMEMIKWYSRR"

BASE COUNT 251 a 182 c 224 g 234 t
ORIGIN

alignment_scores:
  Quality: 609.00      Length: 111
  Ratio: 5.486         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-327-750D-13 x HUOGC ..
  Align seg 1/1 to: HUOGC from: 1 to: 891

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
312 ATGCCAAATATTCACCAAGAAACGAGAGATGGAGCAGCCTATGCAGAA 361

17 nGlyGluGluAspArgProLeuGlyGlyGlyGluGlyHisGlnProAlaG 34
362 TGGAGAGGAGACCGCCCTTTGGGAGGAGGTGAAGCCACCGCCTGCAG 411

34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
412 GAAATCGACGGGACAGGCTCGCGACTTGCCCTAATTTTCGATGGGCC 461

51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
462 ATACCAATAGGCAGATCAATGATGGATGGTGGGAGATGGAGATGATAT 511

67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
512 GGAATATTCATGGAGGAGATGAGAGAAATCAGAAGAAATCTTAGGGAGC 561

84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
562 TGCAGTTGAGGAATTCGCGTATCTTATGGGGAGCTCTCTAATCAC 611

101 HisAspHisHisAspGluPheCysLeuMetPro 111
612 CATGACCATCATGAATTTTGCCTTATGCCT 644

seq_name: gb_pr:AF187064

seq_documentation_block:
  LOCUS AF187064 891 bp mRNA 11-JUN-2000
  DEFINITION Homo sapiens p75NTR-associated cell death executor (NADE) mRNA,
    complete cds.
  ACCESSION AF187064
  VERSION AF187064.1 GI:8452893
  KEYWORDS
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 891)
  AUTHORS Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,
    Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.
  TITLE NADE, a p75NTR-associated cell death executor, is involved in
    signal transduction mediated by the common neurotrophin receptor
    p75NTR
  JOURNAL J. Biol. Chem. 275 (23), 17566-17570 (2000)
```



|||||  
427 TGGAGGAGATGAGAGATCGGAGAAGCTTAGGGAGCTACAGCTGAGA 476  
|||||  
88 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisaspHisH 104  
|||||  
477 AATTGCTACGATCCTTTATGGGGAGCTGTCTAACCCACCACCATCA 526  
|||||

104 sAspGluPheCysLeuMetPro 111  
|||||  
527 TGATGAATTCGCTTATGCT 548  
|||||

seq\_name: gb\_ro:AF097440

seq\_documentation\_block:  
LOCUS AF097440 854 bp mRNA 13-APR-1999  
DEFINITION Mus musculus brain expressed x-linked protein 3 (Bex3) mRNA,  
complete cds.  
ACCESSION AF097440  
VERSION AF097440.1 GI:4580593  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 854)  
Bex1, a gene with increased expression in parthenogenetic embryos,  
Brown, A.L. and Kay, G.F.  
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,  
is a member of a novel gene family on the mouse X chromosome  
Hum. Mol. Genet. 8 (4), 611-619 (1999)  
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)  
MEDLINE 99172070  
REMARK Erratum: [[published erratum appears in Hum Mol Genet 1999  
May;8(5):943]]  
REFERENCE 2 (bases 1 to 854)  
AUTHORS Brown, A.L. and Kay, G.F.  
TITLE Direct Submission  
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of  
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia  
FEATURES  
Location/Qualifiers  
1..854  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/db\_xref="dbEST:AA27375"  
/chromosome="X"  
/map="near Plp"  
/tissue\_type="pooled organs"  
1..854  
/gene="Bex3"  
172..546  
/gene="Bex3"  
/codon\_start=1  
/product="brain expressed X-linked protein 3"  
/protein\_id="AAD24431.1"  
/db\_xref="GI:4580594"  
/translation="MANYHOENEEQPLQGEQDDPVGGEGHQAANNNNNNHNNH  
NNHRRGOARLAPNFRWAIPIRNMNDGLGGDGMFMEMREIRKRLRELQRLNC  
LRILMGEISNHHDDHDECLMP"  
237 a 212 c 228 g 177 t

gene  
CDS  
alignment\_scores:  
Quality: 567.50 Length: 124  
Ratio: 5.113 Gaps: 1  
Percent Similarity: 89.516 Percent Identity: 83.065

alignment\_block:  
US-09-327-750D-13 x AF097440 ..  
Align seg 1/1 to: AF097440 from: 1 to: 854

1 MetalaAsnIleHisGlnAsnGluMetGluGlnProMetGlnAs 17  
|||||

172 ATGCCCAATGTCCACCAGGAAACGAAGAGATGGAGCAGCCCTGCAGAA 221  
|||||  
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyGlyHisGlnProAlaG 34  
|||||  
222 TGGACAGGAAGACCGCCTGTGGAGAGAGGTGAGGCCACACGAGCTGCTG 271  
|||||  
34 lyAsn.....ArgArg 37  
:||||  
272 CAACAACAACAACAACACCAACCAACCAACCAACCAACCAACCAACCA 321  
|||||  
38 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 54  
|||||  
322 GGCCAGGCTCCGCGACTTGGCCCTTAACCTCCGATGGGCCATTCACCAAC 371  
|||||  
54 gGlnIleAsnAspGlyMetGlyGlyAspGlyAspMetGluIlePheM 71  
|||||  
372 GCAGATGAATGACGGGTGGTGGAGATGGAGATGATATGGAAATGTTC 421  
|||||  
71 etGluGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeuArg 87  
|||||  
422 TGGAGGAGATGAGAGATCCGGAGAAAGCTTAGGGAGCTACAGCTGAGA 471  
|||||  
88 AsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisHisaspHisH 104  
|||||  
472 AATTGCTACGATCCTTTATGGGGAGCTGTCTAACCCACCACCATCA 521  
|||||  
104 sAspGluPheCysLeuMetPro 111  
|||||  
522 TGATGAATTCGCTTATGCT 543  
|||||

seq\_name: gb\_ro:AF187065

seq\_documentation\_block:  
LOCUS AF187065 519 bp mRNA 11-JUN-2000  
DEFINITION Rattus norvegicus p75NTR-associated cell death executor (Nade)  
mRNA, complete cds.  
ACCESSION AF187065  
VERSION AF187065.1 GI:8452895  
KEYWORDS Norway rat.  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 519)  
AUTHORS Mukai, J., Hachiya, T., Shoji-Hoshino, S., Kimura, M.T., Nadano, D.,  
Suvanto, P., Hanaoka, T., Li, Y., Irie, S., Greene, L.A. and Sato, T.A.  
TITLE Nade, a p75NTR-associated cell death executor, is involved in  
signal transduction mediated by the common neurotrophin receptor  
p75NTR  
J. Biol. Chem. 275 (23), 17566-17570 (2000)  
JOURNAL 20298829  
MEDLINE 2 (bases 1 to 519)  
REFERENCE Mukai, J., Hachiya, T., Hoshino, S., Kimura, M., Nadano, D.,  
Hanaoka, T., Li, Y., Irie, S. and Sato, T.  
AUTHORS Direct Submission  
TITLE Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia  
JOURNAL University, 630 West 168th St., P&S II-451, New York, NY 10032, USA  
FEATURES  
Location/Qualifiers  
1..519  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
1..519  
/gene="Nade"  
44..406  
/gene="Nade"  
/note="Involved in the common neurotrophin receptor  
p75NTR-mediated signal transduction; Nade"  
/codon\_start=1  
/product="p75NTR-associated cell death executor"  
/protein\_id="AAF75130.1"  
/db\_xref="GI:8452896"

```

MEDLINE 20298829
REFERENCE 2 (bases 1 to 891)
AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
Hanaoka,T., Li,Y., Irie,S. and Sato,T.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES
    source
        1..891
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
    gene
        1..891
    CDS
        312..647
        /gene="NADE"
        /gene="NADE"
        /note="Involved in the common neurotrophin receptor
        p75NTR-mediated signal transduction"
        /codon_start=1
        /product="p75NTR-associated cell death executor"
        /protein_id="AAF75129.1"
        /db_xref="GI:8452894"
        /translation="MANIHQENEEMEQPNQCEEDRPLGGEGHQPAGNRRGOARRLA
        PNFRWAIPIRQINQMGDGDMEIFMEEMREIRKRLRELQLRNLRLIMGELSNHHD
        HHDEFCLMP"
BASE COUNT 251 a 182 c 224 g 234 t
ORIGIN
alignment_scores
    Quality: 609.00 Length: 111
    Ratio: 5.486 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block
    US-09-327-750D-13 x AF187064 ..
    Align seg 1/1 to: AF187064 from: 1 to: 891
1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
312 ATGGCAANTATTCCACGAGAAACGAAGAGATGGAGCGCTATGCAGAA 361
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAlaG 34
362 TGGAGAGCAAGACCGCCCTTTGGGAGGAGGTGAAGGCCACCGCTGCAG 411
34 IyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
412 GAAATCGACGGGACAGCGCTCGCGACTTGCCTTAATTTTCGATGGGCC 461
51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
462 ATACCCCAATAGGCAGATCAATATGGATGGGTGGAGATGGAGATGATAT 511
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
512 GGAATAATTTCATCGAGGAGATGAGAGAAATCAGAAGAAACTTAGGGAGC 561
84 euGlnIleuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
562 TGCAGTTGAGGAAATGTCTCGTATCTTATGGGGAGGCTCTCTAATCAC 611
101 HisAspHisHisAspGluPheCysLeuMetPro 111
612 CATGACCATCATGATGATGATTTTGCCTTATGCCT 644
seq_name: gb_ro:AF187066
seq_documentation_block:
LOCUS AF187066 700 bp mRNA ROD 11-JUN-2000
DEFINITION Mus musculus p75NTR-associated cell death executor (Nade) mRNA,
complete cds.
ACCESSION AF187066

VERSION AF187066.1 GI:8452897
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 700)
Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,
Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.
NADE, a p75NTR-associated cell death executor, is involved in
signal transduction mediated by the common neurotrophin receptor
p75NTR
J. Biol. Chem. 275 (23), 17566-17570 (2000)
REFERENCE 2 (bases 1 to 700)
AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
Hanaoka,T., Li,Y., Irie,S. and Sato,T.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES
    source
        1..700
        /organism="Mus musculus"
        /strain="BALB/c"
        /db_xref="taxon:10090"
    gene
        1..700
    CDS
        177..551
        /gene="Nade"
        /gene="Nade"
        /note="Involved in the common neurotrophin receptor
        p75NTR-mediated signal transduction; NADE"
        /codon_start=1
        /product="p75NTR-associated cell death executor"
        /protein_id="AAF75131.1"
        /db_xref="GI:8452898"
        /translation="MANVHQENEEMEQPLNQGEDRPGGEGHQPAGNRRNNNNHHN
        HNHHRRGOARRLAPNFRWAIPIRQINQMGDGDMEIFMEEMREIRKRLRELQLRNC
        LRLIMGELSNHHDHDEFCLMP"
BASE COUNT 178 a 187 c 203 g 132 t
ORIGIN
alignment_scores
    Quality: 567.50 Length: 124
    Ratio: 5.113 Gaps: 1
    Percent Similarity: 89.516 Percent Identity: 83.065
alignment_block
    US-09-327-750D-13 x AF187066 ..
    Align seg 1/1 to: AF187066 from: 1 to: 700
1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
177 ATGGCAANTATGCCACGAGAAACGAAGAGATGGAGCGCTTGCAGAA 226
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAlaG 34
227 TGGACAGGAAGACCGCCCTGTGGGAGGAGGTGAGGGCCACCGCCCTGCTG 276
34 IyAsn.....ArgArg 37
277 CAACAACAACAACAACCAACCAACCAACCAACCAACCAACCAACCA 326
38 GlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAlaIleProAsnAr 54
327 GGCCAGGCTCGCGACTTGCCTTAACCTTCGATGGCCATTCCTCAACAG 376
54 gGlnIleAsnAspGlyMetGlyGlyAspGlyAspMetGluIlePheM 71
377 GCAGATGAATGACGGGTGGTGGAGATGATATGGAATGTTCATCA 426
71 etGluGluMetArgGluIleArgLysLeuArgGluLeuArg 87
```

/translation="MEQPLQNGQEDRPVGGEGHQPAANNHNNHNNHNNHNNHNNH  
RRGOARRLAPNFRWAIIPNROMNDGLGDDMEMFMEEMREIRKRLRELQRLNCLRLIL  
MGELSNHHDHDEFLMP"  
BASE COUNT 142 a 130 c 136 g 111 t  
ORIGIN

alignment\_scores:  
Quality: 538.50 Length: 125  
Ratio: 5.080 Gaps: 1  
Percent Similarity: 84.800 Percent Identity: 79.200

alignment\_block:  
US-09-327-750D-13 x AF187065 ..

Align seg 1/1 to: AF187065 from: 1 to: 519

6 GlnGluAsnGluMetGluGlnProMetGlnAsnGlyGluGluAspAr 22  
29 CAGGAAACGAGAAATGAGCAGCCCTGCAGATGGACAGGAGACCG 78  
22 gProLeuGlyGlyGluGlyHisGlnProAlaGlyAsn..... 35  
79 CCCCGTGGGAGGAGGTGAGGGCCACCAGCCCTGCTGCAACACACAAACC 128  
36 .....Arg 36  
129 ACAACCATACCAACACCAACCAACCAACCAACCAATCATCACCG 178  
37 ArgGlyGlnAlaArgArgLeuAlaProAsnPhaArgTrpAlaIleProAs 53  
179 AGAGGACAGCCCGCGACTTGCCTTAACCTTCGATGGGCCATTCCTCAA 228  
53 nArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspMetGluIleP 70  
229 CAGGAGATGAATGAGGTGGGTGGGATGGAGATGATATGGAATGT 278  
70 heMetGluMetArgGluIleArgArgLysLeuArgGluLeuGlnLeu 86  
279 TCATGGAGGAGATGAGAGATCCGAGAAAGCTTAGGAGCTGCAGTTG 328  
87 ArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisAspHi 103  
329 AGAAATGTCTGCTATTCTTATGGGGAGCTCTCTAATCACCACGACCA 378  
103 sHisAspGluPheCysLeuMetPro 111  
379 TCACGATGAATTCGCTTATGCCT 403

seq\_name: gb\_sts:G72708

seq\_documentation\_block: 421 bp DNA STS 08-AUG-2001  
LOCUS G72708  
DEFINITION MARC 4953-4954:991939031:1 SCF - porcine spleen Sus scrofa STS  
genomic, sequence tagged site.  
ACCESSION G72708  
VERSION G72708.1 GI:15146738  
KEYWORDS STS.  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 421)  
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L. and  
Keeler,J.W.  
TITLE Single nucleotide polymorphism (SNP) discovery in expressed porcine  
genes  
JOURNAL Unpublished (2001)  
COMMENT  
Contact: Freking BA  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4278

Fax: 402 762 4173  
Email: freking@email.marc.usda.gov  
Primer A: GCAATGGGTGAACACTCTACT  
Primer B: CCAGCAGCAATAGACG  
STS size: 500  
PCR Profile:  
Hotstart: 95 degrees for 15 minutes  
Denature: 95 degrees for 30 seconds  
Anneal: 56 degrees  
Extension: 68 degrees for 2 minutes  
Cycles: 32 to 45

Protocol:  
Template: 50-200 ng genomic DNA  
Primer: each 20 pmoles  
dNTPs: each 88 uM  
Taq Polymerase: 0.25 units (Qiagen HotStar)

Buffer: Commercially supplied Qiagen HotStar buffer

The STS is derived from PCR amplicons generated from genomic DNA, sequenced from each end using the amplification primers. The sequence does not necessarily represent the entire amplicon. Sequence derived from Polyphred was trimmed from each end of each unique contig until five consecutive bases exceeded a quality score threshold of 20, and the next 10 bases averaged a quality score of 20 or greater. Amplicon size was estimated by agarose gel electrophoresis.

FEATURES  
source

location/Qualifiers  
1..421  
/organism="Sus scrofa"  
/strain="white composite, duroc, meishan, minzhu, fengjing, crossbreds"  
/db\_xref="taxon:9823"  
/sex="male and female"  
/clone\_lib="SCF - porcine spleen"  
/dev\_stage="adult"  
/note="Organ: spleen"  
<1..>421

BASE COUNT 130 a 101 c 86 g 102 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 418.00 Length: 77  
Ratio: 5.429 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.701

alignment\_block:  
US-09-327-750D-13 x G72708/rev ..

Align seg 1/1 to reverse of: G72708 from: 1 to: 421

35 AsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPhaArgTrpAlaIle 51  
421 AATAGACCGGACAGCAAGCTGCCGACTTGCCTTAATTCGATGGCCAT 372  
51 eProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspMetG 68  
371 ACCCAATAGCAGATCAATGATGGGATGGGTGGAGATGGAGATGG 322  
68 luIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluLeu 84  
321 AATGTTTCATGGAGAGATGAGAGAAATCAGGAGAAAACCTTAGGAGCTG 272  
85 GlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHisH 101  
271 CAGTTGAGGAATGTCTCGGTATCCTTATGGGGAGCTCTCTAATCACCA 222  
101 sAspHisHisAspGluPheCysLeuMetPro 111  
221 TGACCATCATGATGAATTTTGCCTTATGCCT 191

seq\_name: gb\_pr:HSV351F8

seq\_documentation\_block:

LOGUS HSV351F8 45678 bp DNA PRI 23-NOV-1999  
DEFINITION Human DNA sequence from cosmid V351F8, between markers DXS366 and DXS87 on chromosome X contains ESTs.

ACCESSION 270719  
VERSION 270719.1 GI:1261915  
KEYWORDS X.  
SOURCE human.

#### ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 45678)

AUTHORS Whiteley,M.

TITLE Direct Submission

JOURNAL Submitted (09-APR-1995) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT IMPORTANT: This sequence is the entire insert of clone V351F8. The  
true left end of clone V351F8 is at 1 in this sequence. The true  
right end of clone V351F8 is at 45678.  
V351F8 is from the human chromosome X-specific cosmid library.

#### FEATURES

source

1..45678  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="X"  
/clone="GHC-351F8"  
/clone\_lib="SCCV"  
1..70

repeat\_region

642..930  
/note="Alu repeat: matches 80. .1 of consensus"  
repeat\_region  
1406..5179  
/note="Alu repeat: matches 1. .308 of consensus"  
repeat\_region  
5243..5276  
/note="L1 element fragment"  
repeat\_region  
5279..5567  
/note="17 copies of 2 mer 82 % conserved"

repeat\_region

9975..10369  
/note="Alu repeat: matches 304. .1 of consensus"  
repeat\_region  
10399..12017  
/note="MSTA element fragment"  
repeat\_region  
11318..11978  
/note="MSTAR element fragment"

repeat\_region

11715..11846  
/note="THE1B element fragment"  
repeat\_region  
12074..12170  
/note="MLTIR element fragment"  
repeat\_region  
12227..12486  
/note="MSTA element fragment"  
repeat\_region  
12437..12502  
/note="MSTA element fragment"  
repeat\_region  
12756..13059  
/note="MSTC element fragment"

repeat\_region

14478..14650  
/note="Alu repeat: matches 308. .1 of consensus"  
repeat\_region  
14491..14536  
/note="THE1B element fragment"  
repeat\_region  
14715..14831  
/note="MSTA element fragment"  
repeat\_region  
15099..15333  
/note="THE1B element fragment"  
repeat\_region  
15534..15626  
/note="L1 element fragment"

repeat\_region

15566..15619  
/note="f31 copies of 3 mer 85 % conserved"  
repeat\_region  
16332..16436  
/note="3 copies of 18 mer 98 % conserved"

repeat\_region

/note="L1 element fragment"

repeat\_region  
16728..16763  
/note="9 copies of 4 mer 94 % conserved"  
repeat\_region  
16764..17054  
/note="Alu repeat: matches 308. .1 of consensus"  
repeat\_region  
19420..19449  
/note="15 copies of 2 mer 87 % conserved"  
misc\_feature  
complement(19537..20013)  
/note="match: 3' EST N51315 clone 283089"  
repeat\_region  
20071..20114  
/note="22 copies of 2 mer 98 % conserved"  
repeat\_region  
20073..20112  
/note="10 copies of 4 mer 100 % conserved"  
repeat\_region  
20073..20126  
/note="3 copies of 18 mer 87 % conserved"  
repeat\_region  
23088..23300  
/note="L1 element fragment"  
repeat\_region  
23311..23490  
/note="MLTIA element fragment"  
repeat\_region  
23318..23489  
/note="MLTIB element fragment"  
repeat\_region  
23362..23491  
/note="MLTIC element fragment"  
repeat\_region  
23507..24407  
/note="L1 element fragment"  
repeat\_region  
24376..24462  
/note="MSTC element fragment"  
repeat\_region  
24379..24491  
/note="MSTA element fragment"  
repeat\_region  
24688..24738  
/note="MSTA element fragment"  
repeat\_region  
24810..25016  
/note="L1 element fragment"  
repeat\_region  
25254..25421  
/note="L1 element fragment"  
repeat\_region  
25439..25696  
/note="MER25 element fragment"  
repeat\_region  
26284..26343  
/note="MLTID element fragment"  
repeat\_region  
26428..26491  
/note="MLTIB element fragment"  
repeat\_region  
26465..26518  
/note="MLTIE element fragment"  
repeat\_region  
27446..27667  
/note="L1 element fragment"  
repeat\_region  
28954..29247  
/note="Alu repeat: matches 308. .1 of consensus"  
repeat\_region  
30160..30325  
/note="2 copies of 83 mer 98 % conserved"  
misc\_feature  
32588..32875  
/note="match: 5' EST H68599 clone 239077"  
misc\_feature  
complement(32825..33230)  
/note="match: 3' EST H68239 clone 289077"  
repeat\_region  
35551..35802  
/note="Alu repeat: matches 1. .260 of consensus"  
repeat\_region  
35804..35839  
/note="18 copies of 2 mer 83 % conserved"  
repeat\_region  
38625..38705  
/note="MLTIR element fragment"  
repeat\_region  
38747..38839  
/note="MLTIR element fragment"  
repeat\_region  
38949..39032  
/note="MLTIE element fragment"  
repeat\_region  
38965..39076  
/note="MLTID element fragment"  
BASE COUNT 14260 a 9135 c 9327 g 12956 t  
ORIGIN  
alignment\_scores: 300.00 Length: 114



region 5224. .5555



seq\_name: gb\_sts:G35294

seq\_documentation\_block: 477 bp DNA STS 02-OCT-1997  
 LOCUS G35294 human STS SHGC-37409, sequence tagged site.  
 DEFINITION G35294  
 ACCESSION G35294  
 VERSION G35294.1 GI:2459462  
 KEYWORDS STS; STS sequence; primer; sequence tagged site.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 477)  
 AUTHORS Myers,R.M.  
 JOURNAL Unpublished (1997)

Contact: Richard M. Myers  
 Stanford Human Genome Center (SHGC)  
 Stanford University School of Medicine  
 Department of Genetics, M-344, Stanford, CA 94305, USA  
 Tel: 4157259687  
 Fax: 4157259689  
 Email: myerseshgc.stanford.edu

Primer A: AACATCTTTCATGAAAGTTGATG  
 Primer B: CTTTGGCATCTTCTCGAA  
 STS size: 106  
 PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 15 seconds  
 Annealing: 60 degrees C for 30 seconds  
 Polymerization: 72 degrees C for 23 seconds  
 PCR Cycles: 30  
 Thermal Cycler: Perkin Elmer 9600

Protocol:  
 Template: 25 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 AmpliTaq Gold Polymerase: 0.07 units/ul  
 Total Vol: 5 ul

Buffer:  
 MgCl2: 2.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from N51315  
 -- Washington University/Merck EST sequence.

FEATURES  
 Location/Qualifiers

1..477  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

STS  
 primer\_bind  
 primer\_bind

BASE COUNT 153 a 115 c 79 g 130 t  
 ORIGIN

alignment\_scores:  
 Quality: 249.50 Length: 83  
 Ratio: 3.960 Gaps: 2  
 Percent Similarity: 75.904 Percent Identity: 61.446

alignment\_block:

US-09-327-750D-13 x G35294/rev ..

Align seg 1/1 to reverse of: G35294 from: 1 to: 477

31 GlnProAlaGlyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPh 47

```

:::||||:|||||:::|||||
471 GAGCCTGGAGGAATGTTAAAGGGTTTGGGCTCCACCTGCCCGGGTTT 422
|
47 eArgTrpAlaIleProAsnArgGlnIleAsn.....AspGlyMetGlyG 62
|   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
421 TGGAGAGGATGTGCCCAATAGGCTTGTGCAATAACATTTGATATAGATG 372
|   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
62 lYAspGlyAspAspMetGluIlePheMetGluGluMetArgGluIleArg 78
|   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
371 GAGATGGAGATGATATGGAACGGTTTCATGGAGGAGATGAGAGACTAAGG 322
|   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
79 ArgGlyLeuArgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetG1 95
|   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
321 AGCAAAATTAGGAACTTCAGTTGAGGTACAGCTCGGCATCTCTATAGG 272
|   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
95 yGluLeuSerAsnHisAspHisAspGluPheCysLeuMetPro 111
|   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
271 GGAC...CCTCCTCACCATCATCATCATGAGTTTTCGCTTATGCTT 226
|   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
seq_name: gb_pr:HS714B7

```

seq\_documentation\_block:

LOCUS HS714B7 98274 bp DNA PRI 12-DEC-1999  
 DEFINITION Human DNA sequence from clone CTA-714B7 on chromosome 22q12.2-13.2  
 Contains pseudogene similar to part of COX7B (cytochrome c oxidase  
 subunit VIIb), a novel pseudogene, ESTs, STS and GSSs, complete  
 sequence.  
 ACCESSION Z99755  
 VERSION Z99755.1 GI:3036782  
 KEYWORDS HTG; COX7B; cytochrome c oxidase.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 98274)  
 DIRECT SUBMISSION  
 AUTHOR Connor, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk

COMMENT

On Apr 8, 1998 this sequence version replaced gi:2578134.  
 This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.

This sequence was generated from part of bacterial clone contigs of  
 human chromosome 22, constructed by the Sanger Centre Chromosome 22  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr22  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep CTA-714B7 is  
 from the human BAC library described in U-J. Kim et al. (1996)  
 Genomics 34, 213-218.

VECTOR: pBelOBAC11

This sequence is the entire insert of clone CTA-714B7.

FEATURES

source

Location/Qualifiers  
 1..98274  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /map="q12.2-13.2"  
 /clone="CTA-714B7"



repeat_region	/clone_lib="CIT9785K-A2"	repeat_region	18657..18871
repeat_region	/note="14 copies 2 mer ac 100 conserved"	repeat_region	/note="MIR repeat: matches 35..252 of consensus"
repeat_region	183..258	repeat_region	18927..19230
misc_feature	/note="LMC5 repeat: matches 7840..7913 of consensus"	repeat_region	/note="AluX repeat: matches 1..301 of consensus"
repeat_region	1177..1446	repeat_region	19492..19559
repeat_region	/note="match: GSS: Em:AQ373724"	repeat_region	/note="MIR repeat: matches 2..67 of consensus"
repeat_region	1301..1356	repeat_region	19560..19821
repeat_region	/note="L2 repeat: matches 2693..2750 of consensus"	repeat_region	/note="AluY repeat: matches 34..295 of consensus"
repeat_region	2568..5038	repeat_region	19822..19959
repeat_region	/note="LMC5 repeat: matches 2052..2207 of consensus"	repeat_region	/note="MIR repeat: matches 67..212 of consensus"
repeat_region	5036..5486	repeat_region	19987..20286
repeat_region	/note="LMC5 repeat: matches 2052..2207 of consensus"	repeat_region	/note="AluY repeat: matches 1..300 of consensus"
repeat_region	5496..6885	repeat_region	20316..20512
repeat_region	/note="LMC5 repeat: matches 2052..2207 of consensus"	repeat_region	/note="MIR repeat: matches 82..262 of consensus"
repeat_region	6896..7185	repeat_region	20513..20741
misc_feature	complement(7187..7611)	repeat_region	/note="LMC5 repeat: matches 5970..6224 of consensus"
repeat_region	7369..7594	repeat_region	20769..21074
repeat_region	/note="MIR repeat: matches 34..256 of consensus"	repeat_region	/note="AluSg repeat: matches 1..303 of consensus"
repeat_region	7904..7978	repeat_region	21085..21207
repeat_region	/note="match: GSS: Em:AQ566001"	repeat_region	/note="LMC5 repeat: matches 1776..1898 of consensus"
repeat_region	8086..8300	repeat_region	21212..21391
repeat_region	8154..8565	repeat_region	/note="LMC5 repeat: matches 6..178 of consensus"
repeat_region	/note="match: GSS: Em:AQ59599"	repeat_region	21392..21471
repeat_region	8349..8864	repeat_region	/note="MIR repeat: matches 3..86 of consensus"
repeat_region	/note="AluX repeat: matches 1..280 of consensus"	repeat_region	22146..22441
repeat_region	complement(8897..9290)	repeat_region	/note="AluX repeat: matches 1..296 of consensus"
repeat_region	9235..9361	repeat_region	22526..22610
repeat_region	/note="L2 repeat: matches 2619..2749 of consensus"	repeat_region	/note="LMC5 repeat: matches 267..349 of consensus"
repeat_region	9378..9890	repeat_region	22940..23204
repeat_region	/note="match: GSS: Em:AQ284832"	repeat_region	/note="HERV16 repeat: matches 378..650 of consensus"
repeat_region	10135..10428	repeat_region	23205..23514
repeat_region	/note="AluX repeat: matches 1..294 of consensus"	repeat_region	/note="AluSg repeat: matches 1..311 of consensus"
repeat_region	10429..10563	repeat_region	23919..24194
repeat_region	/note="AluSg repeat: matches 1..135 of consensus"	repeat_region	/note="AluX repeat: matches 36..311 of consensus"
repeat_region	10811..10868	repeat_region	24496..24590
repeat_region	/note="L2 repeat: matches 2648..2705 of consensus"	repeat_region	/note="LMC5 repeat: matches 253..349 of consensus"
repeat_region	10819..11064	repeat_region	24871..24916
repeat_region	/note="MIR repeat: matches 2..255 of consensus"	repeat_region	/note="MIR repeat: matches 102..143 of consensus"
repeat_region	11112..11324	repeat_region	25378..25555
repeat_region	/note="MIR repeat: matches 11..255 of consensus"	repeat_region	/note="L2 repeat: matches 2330..2519 of consensus"
repeat_region	11869..12012	repeat_region	25673..25910
repeat_region	/note="MIR repeat: matches 20..167 of consensus"	repeat_region	/note="MIR repeat: matches 4..255 of consensus"
repeat_region	12925..13223	repeat_region	26722..26798
repeat_region	/note="AluY repeat: matches 1..300 of consensus"	repeat_region	/note="L2 repeat: matches 2572..2649 of consensus"
repeat_region	13290..13495	repeat_region	27040..27107
repeat_region	/note="MER58A repeat: matches 9..214 of consensus"	repeat_region	/note="L2 repeat: matches 2489..2500 of consensus"
repeat_region	13338..13833	repeat_region	27108..27374
repeat_region	/note="match: GSS: Em:AQ568864"	repeat_region	/note="AluY repeat: matches 41..304 of consensus"
repeat_region	13688..13749	repeat_region	27375..27471
repeat_region	/note="MIR repeat: matches 77..139 of consensus"	repeat_region	/note="L2 repeat: matches 2372..2489 of consensus"
repeat_region	13840..13966	repeat_region	27651..27945
repeat_region	/note="MIR repeat: matches 22..147 of consensus"	repeat_region	/note="AluX repeat: matches 1..299 of consensus"
repeat_region	14162..14245	repeat_region	28235..28876
repeat_region	/note="MIR repeat: matches 46..136 of consensus"	repeat_region	/note="L2 repeat: matches 1113..1817 of consensus"
repeat_region	14925..15058	repeat_region	28975..29097
repeat_region	/note="MIR repeat: matches 48..191 of consensus"	repeat_region	/note="MER96 repeat: matches 42..173 of consensus"
repeat_region	15994..16040	repeat_region	29173..29196
repeat_region	/note="MIR repeat: matches 102..144 of consensus"	repeat_region	/note="L2 copies 2 mer aa 100 conserved"
repeat_region	16116..16184	repeat_region	29832..30260
repeat_region	/note="MER5B repeat: matches 3..70 of consensus"	repeat_region	/note="match: GSS: Em:AQ075670"
repeat_region	16897..17005	repeat_region	30392..30597
repeat_region	/note="MIR repeat: matches 9..118 of consensus"	repeat_region	/note="MIR repeat: matches 35..252 of consensus"
repeat_region	17618..17700	repeat_region	30845..31010
repeat_region	/note="MIR repeat: matches 108..192 of consensus"	repeat_region	/note="MER91B repeat: matches 2..162 of consensus"
repeat_region	17748..17912	repeat_region	31056..31094
repeat_region	/note="MIR repeat: matches 73..245 of consensus"	repeat_region	/note="MIR repeat: matches 218..256 of consensus"
		repeat_region	31728..31791
		repeat_region	/note="16 copies 4 mer tggg 76 conserved"
		repeat_region	31858..31952
		repeat_region	/note="MIR repeat: matches 48..153 of consensus"
		repeat_region	complement(33437..33589)
		repeat_region	/note="match: STS: Em:H55731"
		repeat_region	33543..33633

```
/note="MIR repeat: matches 84...165 of consensus"
33785..33966
/note="MER91A repeat: matches 1..165 of consensus"

alignment_scores:
  Quality: 241.50      Length: 67
  Ratio: 4.093         Gaps: 2
Percent Similarity: 88.060 Percent Identity: 77.612

alignment_block:
US-09-327-750D-13 x HS714B7 ..
Align seg 1/1 to: HS714B7 from: 1 to: 98274

43 LeuAlaProAsnPheArgTirAlaIleProAsnArgGlnIleAsnAspG1 59
|||||
62887 TTGGCCCTTAATTTTCAATGGCCATACCAATAGGAGGTCAGATGG 62936
|||||

59 yMetGlyGlyAspGlyAspMetGluIlePheMetGluGluMetArgG 76
|||||
62937 GATGAATAGAGTGAGATGATATGAAATGTTCTATGAGGATGAGAG 62986
|||||

76 luileArgArg-LysLeuArgGluLeuGlnLeuArgAsnCysLeuArgI1 92
|||||
62987 GAATCAGGAGAGAAT...AAGAGCTACAATTTGAGGAATTTCTGTGTAT 63033
|||||

92 eLeuMetGlyGluLeuSerAsnHisHisAspHisHisAspGluPheCys 108
|||||
63034 CCTATATGGAGAG.CHTGTATATCCCATGACCATCATGATGAATTTTGC 63081
|||||

seq_name: gb_pr:AF183416

seq_documentation_block:
LOCUS AF183416 642 bp mRNA PRI 02-SEP-2000
DEFINITION Homo sapiens ovarian granulosa cell 13.0 kDa protein HGR74 homolog
mRNA, complete cds.
ACCESSION AF183416
VERSION AF183416
KEYWORDS AF183416.1 GI:9963770
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Peng,Y., Qian,b., Tu,Y., Xu,S., Han,Z., Fu,G. and Chen,Z.
TITLE A novel gene expressed in human adrenal gland
JOURNAL Unpublished
REFERENCE
AUTHORS Peng,Y., GU,W., Huang,C., Xu,S., Han,Z., Fu,G. and Chen,Z.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1999) Chinese National Human Genome Center at
Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, P. R. China
FEATURES
source
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="adrenal gland"
3..380
/note="HGR74-h protein"
/codon_start=1
/evidence="not_experimental"
/product="ovarian granulosa cell 13.0 kDa protein HGR74
homolog"
/db_xref="GI:9963771"
/translation="MESKEKRAVNSLSMENAOEKEEQVANKGEPLALPLDAGEYC
VPRGNRRFRVROPILQYRWDMMHRLGEPQARMREENMERIGEEVRLMEKLRKQLS
HSLRAVSTDPHPHDDHDECLMP"

BASE COUNT 204 a 118 c 157 g 163 t
ORIGIN
```

```
alignment_scores:
  Quality: 194.50      Length: 113
  Ratio: 2.739         Gaps: 3
Percent Similarity: 62.832 Percent Identity: 43.363

alignment_block:
US-09-327-750D-13 x AF183416 ..
Align seg 1/1 to: AF183416 from: 1 to: 642

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
|||
42 ATGAAATATGCCAACCAAGAAATGAAGAAAGGAGCAAGTTGCTAATAA 91
|||||

17 nGlyGluGlu...AsparGProLeuGlyGlyGlyGluGlyHisGlnProA 33
|||||
92 AGGGAGGCCCTTGGCCCTTGGATGCTGGTGAATACTGTGTGCTTA 141
|||||

33 LaGlyAsnArgArgGly...GlnAlaArgArgLeuAlaProAsnPheArg 48
|||||
142 GAGAAATCTAGCGGTTCCGCTTAGGAGCCCATCTCCAGTATAGA 191
|||||

49 TrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAs 65
|||
192 TGGCATATGATCATAGCTTGGAGACACACAGGAGGATGAGAGAAGA 241
|||||

65 pAspMetGluIlePheMetGluGluMetArgGluIleArgArgLysLeuA 82
|||||
242 GAATATGAAAGGATTTGGGAGAGGTTGACACAGCTGATGGAAAGCTGA 291
|||||

82 rgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSer 98
|||||
292 GGGAAACAGCTTACATAGTCTGGGGCAGTCAGCAGCTGAC...CCC 338
|||||

99 AsnHisHisAspHisHisAspGluPheCysLeuMetPro 111
|||||
339 CCTCACCATGACCATCATGATGAGTTTTCCTTATGCCCC 377
|||||

seq_name: gb_pr:AF220189

seq_documentation_block:
LOCUS AF220189 828 bp mRNA PRI 04-MAY-2000
DEFINITION Homo sapiens uncharacterized hypothalamus protein HBEX2 mRNA,
complete cds.
ACCESSION AF220189
VERSION AF220189.1 GI:7689028
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.
TITLE A novel gene expressed in human hypothalamus
JOURNAL Unpublished
REFERENCE
AUTHORS Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Chinese National Human Genome Center at
Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, P. R. China
FEATURES
source
1..828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="hypothalamus"
167..544
/codon_start=1
/product="uncharacterized hypothalamus protein HBEX2"
/db_xref="AAF67654.1"
/translation="MESKEKRAVNSLSMENAOEKEEQVANKGEPLALPLDAGEYC
VPRGNRRFRVROPILQYRWDMMHRLGEPQARMREENMERIGEEVRLMEKLRKQLS
```

```
HSLRAVSTDPHPHHDHDFCLMP"
BASE COUNT      256 a   170 c   221 g   181 t
ORIGIN

alignment_scores:
  Quality: 194.50      Length: 113
  Ratio: 2.739        Gaps: 3
  Percent Similarity: 62.832  Percent Identity: 43.363

alignment_block:
US-09-327-750D-l3 x AF220189  ..

Align seg 1/1 to: AF220189 from: 1 to: 828

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
   ||| ||| :|||:|||||:||||| :|||: :|||:
206 ATGGAATGCCAACCAAGAAATGAAGAAAGGAGCAAGTGTCTAATAA 255
   ||| ||| :|||:|||||:||||| :|||: :|||:
17 nGlyGluGlu...AspArgProLeuGlyGlyGlyHisGlnProA 33
   :||| ||| :|||:|||||:||||| :|||: :|||:
256 AGGGAGCCCTTGGCCCTTGGATGCTGTAATACTGTGTGCTA 305
   :||| ||| :|||:|||||:||||| :|||: :|||:
33 laGlyAsnArgArgGly...GlnAlaArgLeuAlaProAsnPheArg 48
   ||| ||| :|||:|||||:||||| :|||: :|||:
306 GAGGAATCGTAGCGGTTCGGCTTAGCGAGCCCATCTGCAGTATAGA 355
   ||| ||| :|||:|||||:||||| :|||: :|||:
49 TrpAlaIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAs 65
   ||| ||| :|||:|||||:||||| :|||: :|||:
356 TGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 405
   :||| ||| :|||:|||||:||||| :|||: :|||:
65 pAspMetGluIlePheMetGluMetArgGluIleArgArgLysLeuA 82
   :||| ||| :|||:|||||:||||| :|||: :|||:
406 GAATATGGAAGGATTGGGAGGAGGTGAGACAGCTGATGGAAGAGCTGA 455
   ||| ||| :|||:|||||:||||| :|||: :|||:
82 rgGluLeuGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSer 98
   ||| ||| :|||:|||||:||||| :|||: :|||:
456 GGGAAAGCAGTTGAGTCATAGTCTGGGCGAGTCAGCACTGAC...CCC 502
   ||| ||| :|||:|||||:||||| :|||: :|||:
99 AsnHisHisAspHisHisAspGluPheCysLeuMetPro 111
   ||| ||| :|||:|||||:||||| :|||: :|||:
503 CCTCACCATGACCATGATGATGATGATGATGATGATGATGATGATGATGAT 541
   ||| ||| :|||:|||||:||||| :|||: :|||:
```

**THIS PAGE BLANK (USPTO)**



```
1 MetAlaAsnIleHisGlnGluAsnGluGluMetGluGlnProMetGlnAs 17
|||||
166 ATGCAAAATATTCACAGAGAAACGAAGAGATGGAGCAGCCTATGCAGAA 215
|||||
17 nGlyGluGluAsnArgProLeuGlyGlyGlyGlyGlyGlyHisGlnProAlaG 34
|||||
216 TGGAGAGAGAGAGCGCCCTTTGGAGAGAGGTGAAGGCCACAGCCTGCAG 265
|||||
34 lYAsnArgArgGlyGlnAlaArgAlaArgLeuAlaProAsnPheArgTTPAla 50
|||||
266 GAAATCGACGGGACAGAGCTGCGGACTTGCCCTTAATTTTCGATGGGCC 315
|||||
51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspAspMe 67
|||||
316 ATACCAATAGGCAGATCAATGATGGATGGATGGGTGGAGATGGAGATGATAT 365
|||||
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
366 GGAATATATTCATGGAGAGATGAGAGAAATCAGAAGAAACCTTAGGGAGC 415
|||||
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
416 TGCAGTTGAGGAATGTCTCGGTATCTTATGGGGAGCTCTCTAATCAC 465
|||||
101 HisAspHisHisAspGluPheCysLeuMetPro 111
|||||
466 CATGACCATCATGATCAATTTTGCCTTATGCCT 498
|||||
```

seq\_name: gb\_est2:BG654378

seq\_documentation\_block: 535 bp mRNA EST 05-JUL-2001  
LOCUS BG654378  
DEFINITION ib39e02.y1 HR85 islet Homo sapiens cDNA 5' similar to SW:HG74\_HUMAN  
O00994 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74. ;, mRNA  
sequence.

ACCESSION BG654378  
VERSION BG654378.1 GI:13791787  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 535)

REFERENCE  
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
Lemishka,I., Seearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,  
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas  
,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,  
Jackson,Y. and Bowers,Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)

TITLE  
JOURNAL  
COMMENT Other\_ESTs: ib39e02.x1  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@bioh.p.harvard.edu  
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)

Seq primer: -40RP from Gibco  
High quality sequence stop: 422.  
Location/Qualifiers  
1..535  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HR85 islet"  
/tissue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"

FEATURES  
source

/note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1:  
NotI; Site\_2: XhoI; cDNA made by oligo-dT priming. -1kb. 5'  
Size selected on agarose gel. Average insert size -1kb. 5'  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permutt Lab), Washington University  
School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:  
314-362-1916, Fax: 314-747-2692."

BASE COUNT 153 a 121 c 160 g 101 t  
ORIGIN

alignment\_scores:  
Quality: 609.00 Length: 111  
Ratio: 5.486 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-327-750D-13 x BG654378

Align seg 1/1 to: BG654378 from: 1 to: 535  
1 MetAlaAsnIleHisGlnGluAsnGluGluMetGluGlnProMetGlnAs 17  
|||||  
177 ATGCAAAATATTCACAGAGAAACGAAGAGATGGAGCAGCCTATGCAGAA 226  
|||||  
17 nGlyGluGluAsnArgProLeuGlyGlyGlyGlyGlyGlyHisGlnProAlaG 34  
|||||  
227 TGGAGAGAGAGCGCCCTTTGGAGAGAGGTGAAGGCCACAGCCTGCAG 276  
|||||  
34 lYAsnArgArgGlyGlnAlaArgAlaArgLeuAlaProAsnPheArgTTPAla 50  
|||||  
277 GAAATCAGCGGGACAGAGCTCGCCGACTTGCCCTTAATTTTCGATGGGCC 326  
|||||  
51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspAspMe 67  
|||||  
327 ATACCAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGATGATAT 376  
|||||  
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84  
|||||  
377 GGAATATATTCATGGAGAGATGAGAGAAATCAGAAGAAACCTTAGGGAGC 426  
|||||  
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100  
|||||  
427 TGCAGTTGAGGAATGTCTCGGTATCTTATGGGGAGCTCTCTAATCAC 476  
|||||  
101 HisAspHisHisAspGluPheCysLeuMetPro 111  
|||||  
477 CATGACCATCATGATCAATTTTGCCTTATGCCT 509  
|||||

seq\_name: gb\_est1:AW512400

seq\_documentation\_block:  
LOCUS AW512400 612 bp mRNA EST 03-MAR-2000  
DEFINITION xx74b09.x1 NCI-CGAP\_Lym12 Homo sapiens cDNA clone IMAGE:2849369 3'  
Similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74  
(HUMAN);, mRNA sequence.

ACCESSION AW512400  
VERSION AW512400.1 GI:7150478  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 612)  
REFERENCE  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Life Technologies catalog #: 11547-015

DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 image.llnl.gov/image/html/iresources.shtml

Seq primer: -40UP from Gibco  
 High quality sequence stop: 422.

#### FEATURES

Location/Qualifiers  
 1..612  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2849369"  
 /clone\_lib="NCI-CGAP\_Lym12"  
 /tissue\_type="lymphoma, follicular mixed small and large  
 cell"  
 /lab\_host="DH10B"  
 /note="Organ: lymph node; Vector: pCMV-SPORT6; Site.1:  
 SalI; Site.2: NotI; Cloned unidirectionally. Primer:  
 Oligo dt. Average insert size 1.25 kb. Life Technologies  
 catalog #: 11547-015"  
 catalog # 11547-015"  
 BASE COUNT 169 a 145 c 120 g 178 t  
 ORIGIN

alignment\_scores:  
 Quality: 609.00 Length: 111  
 Ratio: 5.486 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-327-750D-13 x AW512400/rev ..

Align seg 1/1 to reverse of: AW512400 from: 1 to: 612

```

1 MetAlaAsnIleHisGlnGluAsnGluMetClnProMetGlnAs 17
|||||
581 ATGCCAAATATTACCCAGGAGAAACGAGAGATGGAGCAGCTATGCAGAA 532

17 nGlyGluAspArgProLeuGlyGlyGlyGluGlyHisGlnProAla 34
|||||
531 TGGAGAGAGACCGCCCTTTGGGAGAGGTGAAGCCACCAGCCTGCAG 482

34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
|||||
481 GAAATCGACGGGACAGCGCTCGCGACTTGCCTTAATTTTCGATGGCC 432

51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
|||||
431 ATACCCCAATAGGCAGATCAATGATGGGATGGGTGGAGATGGAGATGAT 382

67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGlu 84
|||||
381 GGAATATTTCATGGAGAGATGAGAGAAATCAGAGAAACTTAGGAGC 332

84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
331 TGCAGTTGAGGAATTTGCTCGTATCCTTTATGGGGAGCTCTCTAATCAC 282

101 HisAspHisHisAspGluPheCysLeuMetPro 111
|||||
281 CATGACCATCATGATGAATTTTGCCTTATGCCT 249
```

seq\_name: gb\_est1:A1193112

#### seq\_documentation\_block:

LOCUS A1193112 615 bp mRNA EST 29-OCT-1998  
 DEFINITION qe69g09.x1 Soares\_fetal\_lung\_NbHL19W Homo sapiens cDNA clone  
 IMAGE:1744288 3' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0  
 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.

ACCESSION A1193112  
 VERSION A1193112.1 GI:3744321  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

#### REFERENCE AUTHORS TITLE JOURNAL COMMENT

1 (bases 1 to 615)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-re@mail.nih.gov

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 845 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 445.

#### FEATURES

Location/Qualifiers  
 1..615  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1744288"  
 /clone\_lib="Soares\_fetal\_lung\_NbHL19W"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: lung; Vector: pT73D (Pharmacia) with a  
 modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTTTTTTTTT-3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M.Fatima Bonaldo. This library was constructed  
 from the same fetus as the fetal heart library, Soares  
 fetal heart NbHL19W."

BASE COUNT 169 a 145 c 121 g 180 t  
 ORIGIN

#### alignment\_scores:

Quality: 609.00 Length: 111  
 Ratio: 5.486 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-327-750D-13 x A1193112/rev ..

Align seg 1/1 to reverse of: A1193112 from: 1 to: 615

```

1 MetAlaAsnIleHisGlnGluAsnGluMetClnProMetGlnAs 17
|||||
584 ATGCCAAATATTACCCAGGAGAAACGAGAGATGGAGCAGCTATGCAGAA 535

17 nGlyGluClnAspArgProLeuGlyGlyGlyGlyGlyHisGlnProAla 34
|||||
534 TGGAGAGAGACCGCCCTTTGGGAGGAGGTCAAGCCACCAGCCTGCAG 485

34 lyAsnArgArgGlyGluAlaArgArgLeuAlaProAsnPheArgTrpAla 50
|||||
484 GAAATCGACGGGACAGCGCTCGCGACTTGCCTTAATTTTCGATGGGCC 435

51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
|||||
434 ATACCCCAATAGGCAGATCAATGATGGGATGGGTGGAGATGGAGATGAT 385

67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGlu 84
|||||
384 GGAATATTTCATGGAGAGATGAGAGAAATCAGAGAAACTTAGGAGC 335

84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
334 TGCAGTTGAGGAATTTGCTCGTATCCTTTATGGGGAGCTCTCTAATCAC 285

101 HisAspHisHisAspGluPheCysLeuMetPro 111
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 615)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-re@mail.nih.gov

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 845 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 445.

#### FEATURES

Location/Qualifiers  
 1..615  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1744288"  
 /clone\_lib="Soares\_fetal\_lung\_NbHL19W"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: lung; Vector: pT73D (Pharmacia) with a  
 modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTTTTTTTTT-3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M.Fatima Bonaldo. This library was constructed  
 from the same fetus as the fetal heart library, Soares  
 fetal heart NbHL19W."

BASE COUNT 169 a 145 c 121 g 180 t  
 ORIGIN

#### alignment\_scores:

Quality: 609.00 Length: 111  
 Ratio: 5.486 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-327-750D-13 x A1193112/rev ..

Align seg 1/1 to reverse of: A1193112 from: 1 to: 615

```

1 MetAlaAsnIleHisGlnGluAsnGluMetClnProMetGlnAs 17
|||||
584 ATGCCAAATATTACCCAGGAGAAACGAGAGATGGAGCAGCTATGCAGAA 535

17 nGlyGluClnAspArgProLeuGlyGlyGlyGlyGlyHisGlnProAla 34
|||||
534 TGGAGAGAGACCGCCCTTTGGGAGGAGGTCAAGCCACCAGCCTGCAG 485

34 lyAsnArgArgGlyGluAlaArgArgLeuAlaProAsnPheArgTrpAla 50
|||||
484 GAAATCGACGGGACAGCGCTCGCGACTTGCCTTAATTTTCGATGGGCC 435

51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
|||||
434 ATACCCCAATAGGCAGATCAATGATGGGATGGGTGGAGATGGAGATGAT 385

67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGlu 84
|||||
384 GGAATATTTCATGGAGAGATGAGAGAAATCAGAGAAACTTAGGAGC 335

84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
334 TGCAGTTGAGGAATTTGCTCGTATCCTTTATGGGGAGCTCTCTAATCAC 285

101 HisAspHisHisAspGluPheCysLeuMetPro 111
```

```

|||||
284 CATGACCATCATGATGAATTTTGGCTTATGCCT 252
seq_name: gb_est1:AA576958

seq_documentation_block: 620 bp mRNA 12-SEP-1997
LOCUS AA576958 EST
DEFINITION nm82d04.s1 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:1074727 3',
similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74
(HUMAN);, mRNA sequence.
ACCESSION AA576958
VERSION AA576958.1 GI:2354432
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 620)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 725 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 339.
FEATURES
source
1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1074727"
/clone_lib="NCI_CGAP_Co9"
/tissue_type="colon tumor RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo (Soares4
)."
BASE COUNT 169 a 144 c 116 g 191 t
ORIGIN

alignment_scores:
Quality: 609.00 Length: 111
Ratio: 5.486 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-327-750D-13 x AA576958/rev ..
Align seg 1/1 to reverse of: AA576958 from: 1 to: 620

1 MetAlaAsnIleHisGlnGluAsnGluGluMetGluGlnProMetGlnAs 17
|||||
602 ATGGCAAAATATTCCACGAGAAACGAGAGATGGAGCAGCTATGCAGAA 553
|||||
17 nGlyGluGluAsnArgProLeuGlyGlyGlyGluGlyHisGlnProAlaG 34
|||||
552 TGGAGAGGAGACCGCCCTTTGGGAGGAGGTGAAGCCACGACCTGCAG 503
|||||
34 LyAsnArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
|||||
502 GAAATCGACGGGACAGCTCGCCGACTTGCCCTTAATTTTCGATGGGCC 453
|||||
51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyGlyAspGlyAspPme 67
|||||
452 ATACCAATAGGCAGATCAATGATGGTGGTGGTGGATGAGATGATAT 403
|||||
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
402 GGAATATTTCATGGAGGAGATGAGAGAAATCAGAGAAACTTAGGGAGC 353
|||||
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
352 TGCAGTTGAGGAATGTCTCGTATCCTTATGGGGAGCTCTCTAATCAC 303
|||||
101 HisAspHisHisAspGluPheCysLeuMetPro 111
|||||
302 CATGACCATCATGATGAATTTTGCCTTATGCCT 270
|||||
seq_name: gb_est1:AI929106

seq_documentation_block: 662 bp mRNA 23-AUG-1999
LOCUS AI929106 EST
DEFINITION au65b10.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone
IMAGE:2519611 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION AI929106
VERSION AI929106.1 GI:5665070
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 662)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisler,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 437.
FEATURES
Location/Qualifiers
1..662
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2519611"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: Ssti; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGGATCTTAATTAATTAATCCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGCTCGAGTCTTTTTTTTTTTTTTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length

```



enrichment and has not undergone amplification. Library was constructed by Dr. Claudio Schneider (LNCIB-Area Science Park, Trieste, Italy). "

BASE COUNT 178 a 143 c 169 g 170 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 609.00 Length: 111  
Ratio: 5.486 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-13 x AI929106 ..

Align seg 1/1 to: AI929106 from: 1 to: 662

```

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
|||||
96 ATGCCAAATATTCCACGAGAAACGAAGAGATGGAGCAGCCTATGCAGAA 145

17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAlaG 34
|||||
146 TGGAGAGGAAGACCGCCCTTTGGGAGGAGGTGAAGGCCACCCAGCCTGCAG 195

34 lYAsnArgArgGlyGlnAlaArgLeuAlaProAsnPheArgTTPAla 50
|||||
196 GAAATCGACGGGACAGGCTCGCGACTTGCCTTAATTTTCGATGGGCC 245

51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
|||||
246 ATACCCAAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGATGATAT 295

67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
296 GGAATATTATCATGGAGGAGATGAGAGAAATCAGAAGAAACATTAGGAGC 345

84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
346 TGCAGTTGAGGATTTCTGCGTATCTTATGGGGAGGCTCTCTAATCAC 395

101 HisAspHisHisAspGluPheCysLeuMetPro 111
|||||
396 CATGACCATCATGATGAATTTTGCCTTATGCCT 428

```

seq\_name: gb\_est2:BG745809

seq\_documentation\_block:

LOCUS BG745809 674 bp mRNA EST 15-MAY-2001  
DEFINITION 602723951T1 NIH\_MGC\_113 Homo sapiens cDNA clone IMAGE:4850293 3', mRNA sequence.  
ACCESSION BG745809  
VERSION BG745809.1 GI:14056462  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 674)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Mark Watson  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: L1691691 row: e column: 14  
High quality sequence stop: 674.

REFERENCE

AUTHORS  
TITLE  
JOURNAL  
COMMENT

FEATURES  
Source

Location/Qualifiers

```

1..674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4850293"
/lab_host="NIH_MGC_113"
/note="Organ: spleen; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

```

BASE COUNT 167 a 186 c 142 g 179 t  
ORIGIN

alignment\_scores:  
Quality: 609.00 Length: 111  
Ratio: 5.486 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-13 x BG745809/rev ..

Align seg 1/1 to reverse of: BG745809 from: 1 to: 674

```

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
|||||
529 ATGCCAAATATTCCACGAGAAACGAAGAGATGGAGCAGCCTATGCAGAA 480

17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyHisGlnProAlaG 34
|||||
479 TGGAGAGGAAGACCGCCCTTTGGGAGGAGGTGAAGGCCACCCAGCCTGCAG 430

34 lYAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTTPAla 50
|||||
429 GAAATCGACGGGACAGGCTCGCGACTTGCCTTAATTTTCGATGGGCC 380

51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
|||||
379 ATACCCAAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGATGATAT 330

67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
329 GGAATATTATCATGGAGGAGATGAGAGAAATCAGAAGAAACATTAGGAGC 280

84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
279 TGCAGTTGAGGATTTCTGCGTATCTTATGGGGAGGCTCTCTAATCAC 230

101 HisAspHisHisAspGluPheCysLeuMetPro 111
|||||
229 CATGACCATCATGATGAATTTTGCCTTATGCCT 197

```

seq\_name: gb\_est1:AV701506

seq\_documentation\_block:

LOCUS AV701506 720 bp mRNA EST 08-OCT-2000  
DEFINITION AV701506 ADB Homo sapiens cDNA clone ADBBBH12 5', mRNA sequence.  
ACCESSION AV701506  
VERSION AV701506.1 GI:10717836  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 720)  
Peng, Y., Song, H., Huang, Q., Huang, C., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 586)
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/hcicgap">http://www.ncbi.nlm.nih.gov/hcicgap</a> .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:CGAPs-r@mail.nih.gov">CGAPs-r@mail.nih.gov</a> Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> Seq primer: -400P from Gibco High quality sequence set: 464. Location/Qualifiers 1..586
FEATURES	1..586

```

/od_xref="taxon:9806"
/clone="IMAGE:3477674"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrothecoma"
/lab_host="DH10B (phage-resistant)"
/notes=Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAAGTCGAGCGCGCGGCACATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      166 a   143 c   112 g   164 t      1 others
ORIGIN

alignment_scores:
    Quality:      600.00      Length:      111
    Ratio:        5.455              Gaps:      0
Percent Similarity: 99.099      Percent Identity: 99.099

alignment_block:
US-09-327-750D-13 x BF058872/rev ..

Align seg 1/1 to reverse of: BF058872 from: 1 to: 586

1 MetAlaAsnIleHisGlnGluAsnGluGluMetGluGlnPrometGlnAs 17
586 ATGCGAAATATTCCACGAGAAAACGACAGATGGCAGCAGCNTATGCGAGA 537
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyGlyHisGlnProAlaG 34
536 TGGAGAGGAAGACCGCCCTTTGGAGAGGAGGTGAAGGCCACCGAGCTGCG 487
34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
486 GAAATCGAGCGGGACAGCGCTCCCGACTTGGCCCTTAATTTTCGATGGGCC 437
51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPhe 67
436 ATACCAATAGCAGATCAATGATGGGATGGTGGAGATGGAGATGATAT 387
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
386 GGAAATATTTCATGGAGGAGATGAGAGAAATCAGAAGAAACTTAGGAGC 337

```

```

84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
336 TGCAGTTGAGGAATGCTCGGTATCCTTATGGGGAGCTCTCTAATCAC 287
|||||
101 HisAspHisHisAspGluPheCysLeuMetPro 111
|||||
286 CATGACCATCATGATGAATTTGGCCTTATGCCT 254
|||||
seq_name: gb_est2:BG482054

seq_documentation_block:
LOCUS BG482054 557 bp mRNA EST 21-MAR-2001
DEFINITION 602527934F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4651353 5',
mRNA sequence.
ACCESSION BG482054
VERSION BG482054.1 GI:13414333
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 557)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1434 row: d column: 10
High quality sequence stop: 556.
Location/Qualifiers
1..557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4651353"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 157 a 126 c 166 g 108 t

ORIGIN

alignment\_scores:

Quality: 597.00 Length: 111  
Ratio: 5.427 Gaps: 0  
Percent Similarity: 99.099 Percent Identity: 99.099

alignment\_block:

US-09-327-750D-13 x BG482054 ..

Align seg 1/1 to: BG482054 from: 1 to: 557

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17

|||||

185 ATGCCAATATTCCACGAGAAACGAGAGATGGAGCAGCCTATGCAGAA 234

|||||

17 nGlyGluGluAspArgProLeuGlyGlyGlyHisGlnProAlaG 34

|||||

235 TGGAGAGGAACGCCGCTTTGGGAGGAGGTGAAGGCCACCGCCTGCAG 284

```

34 lYAsnArgArgGlyGlnAlaArgLeuAlaProAsnPheArgTtpAla 50
|||||
285 GAAATCGACGGGACAGGCTCGCCGACTTCCCTTATTTTCGATGGGCC 334
|||||
51 lIleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
|||||
335 ATACCCAATAGGCAGATCAATGATGGGATGGGTGGAGATGGAGATGAT 384
|||||
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
385 GGAATATTTCATGGAGAGATGAGAGAAATCAGAGAAACTTAGGGAGC 434
|||||
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
435 TGCAGTTGAGGAATGCTCGGTATCCTTATGGGGAGCTCTCTAATCAC 484
|||||
101 HisAspHisHisAspGluPheCysLeuMetPro 111
|||||
485 CATGACCATCATGATGAATTTGGCCTTATGCCT 517
|||||
seq_name: gb_est2:BG705685
```

seq\_documentation\_block:

LOCUS BG705685 567 bp mRNA EST 07-MAY-2001  
DEFINITION 602668910F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4791506 5',  
mRNA sequence.

ACCESSION BG705685

VERSION BG705685.1 GI:13980274

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 567)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM10568 row: d column: 03

High quality sequence stop: 567.

FEATURES

source

Location/Qualifiers

1..567

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4791506"

/clone\_lib="NIH\_MGC\_96"

/tissue\_type="hypothalamus"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site:1: BamHI; Site:2: Sali-XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.3 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NHGRI), National  
Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 160 a 130 c 168 g 109 t

ORIGIN

alignment\_scores:

Quality: 597.00

Length: 111

Ratio: 5.427 Gaps: 0  
Percent Similarity: 99.099 Percent Identity: 99.099

## alignment\_block:

US-09-327-750D-13 x BG705685

Align seg 1/1 to: BG705685 from: 1 to: 567

```

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
|||||
194 ATGCCAAATATTCCACGAGAAACGAAGAGATGGAGCAGCCTATGCAGAA 243
|||||
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyGlyGlyHisGlnProAlaG 34
|||||
244 TGGAGAGGAAGACCGCCCTTTGGGAGGAGGTGAAGGCCACCGCCTGCAG 293
|||||
34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrrPala 50
|||||
294 GAAATCGACGGGACGAGCTCGCGACTTGCCTTAATTTTCGATGGGCC 343
|||||
51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
|||||
344 ATACCCAAATAGGCAGATCAATGATGGGATGGGTGGAGATGGAGATGATAT 393
|||||
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
394 GGAATATTTCATGGAGGAGATGAGAGAAATCAGAAGAAACTTAGGGAGC 443
|||||
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
444 TGCAGTTGAGGAATTTGTCGATATCCTTATGGGGAGGCTCTCTAATCAC 493
|||||
101 HisAspHisAspGluPheCysLeuMetPro 111
|||||
494 CATGACCATCATGATGAATTTGGCCTTATGCCT 526
|||||

```

seq\_name: gb\_est2:BE896163

```

seq_documentation_block:
LOCUS BE896163 676 bp mRNA EST 20-OCT-2000
DEFINITION 601438756F1 NIH_MGC-72 Homo sapiens cDNA clone IMAGE:3923941 5',
mRNA sequence.
ACCESSION BE896163
VERSION BE896163.1 GI:10360109
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 676)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: ATCC/DCTD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9760 row: 0 column: 14
High quality sequence stop: 671.
Location/Qualifiers
1. .676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3923941"
/clone_lib="NIH_MGC-72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;

```

## FEATURES

source

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 179 a 147 c 175 g 175 t  
ORIGIN

alignment\_scores:  
Quality: 597.00 Length: 111  
Ratio: 5.427 Gaps: 0  
Percent Similarity: 99.099 Percent Identity: 99.099

## alignment\_block:

US-09-327-750D-13 x BE896163

Align seg 1/1 to: BE896163 from: 1 to: 676

```

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
|||||
107 ATGCCAAATATTCCACGAGAAACGAAGAGATGGAGCAGCCTATGCAGAA 156
|||||
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyGlyGlyHisGlnProAlaG 34
|||||
157 TGGAGAGGAAGACCGCCCTTTGGGAGGAGGTGAAGGCCACCGCCTGCAG 206
|||||
34 lyAsnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrrPala 50
|||||
207 GAAATCGACGGGACGAGCTCGCGACTTGCCTTAATTTTCGATGGGCC 256
|||||
51 IleProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspPme 67
|||||
257 ATACCCAAATAGGCAGATCAATGATGGGATGGGTGGAGATGGAGATGATAT 306
|||||
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
307 GGAATATTTCATGGAGGAGATGAGAGAAATCAGAAGAAACTTAGGGAGC 356
|||||
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGluLeuSerAsnHis 100
|||||
357 TGCAGTTGAGGAATTTGTCGATATCCTTATGGGGAGGCTCTCTAATCAC 406
|||||
101 HisAspHisAspGluPheCysLeuMetPro 111
|||||
407 CATGACCATCATGATGAATTTGGCCTTATGCCT 439
|||||

```

seq\_name: gb\_est2:BG993540

```

seq_documentation_block:
LOCUS BG993540 550 bp mRNA EST 13-JUN-2001
DEFINITION PM9-HT0911-080201-012-h10 HT0911 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG993540
VERSION BG993540.1 GI:14397610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 550)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

```

## TITLE

sequence tags

JOURNAL

MEDLINE

COMMENT

Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&t2=PM0-HT0911-080201-012-h10&t3=2001-02-08&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 550.

## FEATURES

source  
 1..550  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HT0911"  
 /dev\_stage="Adult"  
 /note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 154 a 113 c 137 g 146 t  
 ORIGIN

alignment\_scores:  
 Quality: 596.00 Length: 112  
 Ratio: 5.369 Gaps: 1  
 Percent Similarity: 99.107 Percent Identity: 99.107

## alignment\_block:

US-09-327-750D-13 x BG993540 ..  
 Align seg 1/1 to: BG993540 from: 1 to: 550

```

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
|||||
48 ATGGCAAAATATTCCACGAGAAACGAAGAGATGGAGCAGCCTATGCAGAA 97
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyGlyGlyHisGlnProAlaG 34
|||||
98 TGGAGAGAAAGACCGCCCTTTGGGAGGAGGTGAAGGCCACAGCCTGCAG 147
34 lYasnArgArgGlyGlnAlaArgArgLeuAla.ProAsnPheArgTrpAl 50
|||||
148 GAAATCAGCGGGACAGGCTCGCGACTTGCCTTAATTTTCGATGGGC 197
50 aileProAsnArgGlnIleAsnAspGlyMetGlyGlyAspGlyAspAspM 67
|||||
198 CATACCAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGATGATA 247
67 etGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGlu 83
|||||
248 TGGAAATATTTCATGGAGGAGATGAGAAATCAGAAGAAACTTAGGGAG 297
84 LeuGlnLeuArgAsnCysLeuArgIleLeuMetClyGlyLeuSerAsnHi 100
|||||
298 CTGCAGTTGAGGAATTCCTCGCTATCCTTATGGGGAGGCTCTTAATCA 347
100 sHisAspHisHisAspGluPheCysLeuMetPro 111
|||||
348 CCATGACCATCATGATGAATTTGCCTTATGCCT 381
```

seq\_name: gb\_est1:BE252272

seq\_documentation\_block:  
 LOCUS BE252272 559 bp mRNA EST 13-JUL-2000  
 DEFINITION 601114115F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:3354819 5',  
 mRNA sequence.  
 ACCESSION BE252272  
 VERSION BE252272.1 GI:9122409  
 KEYWORDS EST.

## SOURCE

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (Bases 1 to 559)  
 AUTHORS NIH-MGC http://www.ncbi.nih.gov/.  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM157 row: n column: 04  
 High quality sequence stop: 559.

## FEATURES

source  
 1..559  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3354819"  
 /clone\_lib="NIH\_MGC\_16"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCAGCAG(G). Library constructed by Ling Hong  
 in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 157 a 126 c 166 g 110 t  
 ORIGIN

alignment\_scores:  
 Quality: 591.00 Length: 111  
 Ratio: 5.422 Gaps: 0  
 Percent Similarity: 98.198 Percent Identity: 98.198

## alignment\_block:

US-09-327-750D-13 x BE252272 ..

Align seg 1/1 to: BE252272 from: 1 to: 559

```

1 MetAlaAsnIleHisGlnGluAsnGluMetGluGlnProMetGlnAs 17
|||||
177 ATGGCAAAATATTCCACGAGAAACGAAGAGATGGAGCAGCCTATGCAGAA 226
17 nGlyGluGluAspArgProLeuGlyGlyGlyGlyGlyHisGlnProAlaG 34
|||||
227 TGGAGAGAAAGACCGCCCTTTGGGAGGAGGTGAAGGCCACAGCCTGCAG 276
34 lYasnArgArgGlyGlnAlaArgArgLeuAlaProAsnPheArgTrpAla 50
|||||
277 GAAATCAGCGGGACAGGCTCGCGACTTGCCTTAATTTTCGATGGGCC 326
51 ileProAsnArgGlnIleAsnAspGlyMetClyGlyAspGlyAspAspMe 67
|||||
327 ATACCAATAGGCAGATCAATGATGGATGGGTGGAGATGGAGATGATAT 376
67 tGluIlePheMetGluGluMetArgGluIleArgArgLysLeuArgGluL 84
|||||
377 GAAATATTTCATGGAGGAGATGAGAAATCAGAAGAAACTTAGGGAGC 426
84 euGlnLeuArgAsnCysLeuArgIleLeuMetGlyGlyGlyLeuSerAsnHis 100
|||||
427 TGCAGTTGAGGAATTCCTCGTATCCTTATGGGGAGGACTCTCTAATCAC 476
101 HisAspHisHisAspGluPheCysLeuMetPro 111
```

```
|||||CATGACCATCATGATGAATTTGGCTTATGCCT 509
seq_name: gb_est1:AI928835
seq_documentation_block: 626 bp mRNA EST 23-AUG-1999
LOCUS AI928835 au65b10.x1 Schneider fetal brain 00004 Homo sapiens CDNA clone
DEFINITION IMAGE:2519611.3', similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
KD PROTEIN HGR74 (HUMAN); mRNA sequence.
ACCESSION AI928835
VERSION AI928835.1 GI:5664828
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 626)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,X., Wylie,T., Waterston,R. and Willson,R.
WASHU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LILNL; contact the
IMAGE Consortium (info@image.lilnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 475.
FEATURES
Location/Qualifiers
1..626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2519611"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site1: SstI; Site2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCCTCCCTCCCTCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGACTCGAGTGTGTTTGTGTTTGTGTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."
```

BASE COUNT 175 a 144 c 122 g 184 t 1 others

alignment\_scores: Quality: 589.00 Length: 111  
Ratio: 5.355 Gaps: 0  
Percent Similarity: 99.099 Percent Identity: 97.297

alignment\_block:  
US-09-327-750D-13 x AI928835/rev

Align seg 1/1 to reverse of: AI928835 from: 1 to: 626

1 MetAlaAsnIleHisGlnGluAsnGluGluMetGluGlnProMetGlnAs 17

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2002, 14:17:05 ; Search time 1319.57 Seconds  
(without alignments)  
8751.355 Million cell updates/sec

Title: US-09-327-750D-28

Perfect score: 700

Sequence: 1 acagcgctgtgccagcgc.....ggggctgtgtgccagtga 700

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.om.\*

20: em.or.\*

21: em.ov.\*

22: em.pat.\*

23: em.ph.\*

24: em.pl.\*

25: em.ro.\*

26: em.sts.\*

27: em.sy.\*

28: em.un.\*

29: em.vi.\*

30: em.htgo\_hum.\*

31: em.htgo\_inv.\*

32: em.htgo\_rod.\*

33: em.htg\_hum.\*

34: em.htg\_inv.\*

35: em.htg\_rod.\*

36: em.htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	692	98.9	700	10	AF187066	AF187066 Mus muscu
2	676	96.6	854	10	AF097440	AF097440 Mus muscu
3	355.8	50.8	519	10	AF187065	AF187065 Rattus no
4	337.8	48.3	793	9	BC003190	BC003190 Homo sapi
5	307.4	43.9	891	9	HU000319	M38188 Human unkno
6	307.4	43.9	891	9	AF187064	AF187064 Homo sapi
7	240.6	34.4	421	11	G72708	G72708 MARC 4953-4
8	154.8	22.1	785	10	AF097439	AF097439 Mus muscu
9	154.6	22.1	98274	9	HS71487	Z99755 Human DNA s
10	101.4	14.5	477	11	G35294	G35294 human STS S
11	101.4	14.5	45678	9	HSV351F8	Z70719 Human DNA s
12	80.2	11.5	1229	9	AK000959	AK000959 Homo sapi
13	80.2	11.5	1364	6	AX100231	AX100231 Sequence
14	80.2	11.5	69648	9	HS635G19	AL035494 Human DNA
15	77.4	11.1	244	11	GL3550	GL3550 human STS S
16	67.6	9.7	99	10	MUSCEREBL	LI0905 Mus musculus
17	59.8	8.5	7218	6	I66494	I66494 Sequence 14
18	58	8.3	642	9	AF183416	AF183416 Homo sapi
19	58	8.3	828	9	AF220189	AF220189 Homo sapi
20	58	8.3	35714	9	HS198P4	AL008708 Human DNA
21	57.2	8.2	43952	9	HSU105G4	Z92846 Human DNA s
22	56.4	8.1	791	9	AF237783	AF237783 Homo sapi
23	55.8	8.0	744	9	AF251053	AF251053 Homo sapi
24	55.8	8.0	898	6	AX078272	AX078272 Sequence
25	55.8	8.0	31321	9	HSV870H8	Z70233 Human DNA s
26	55.8	8.0	41029	2	HSU80B1	AL022169 Homo sapi
27	54.4	7.8	504	11	G24641	G24641 human STS W
28	54.2	7.7	40584	9	AL133348	AL133348 Human DNA
29	54	7.7	835	10	AF097438	AF097438 Mus muscu
30	54	7.7	838	10	AF051347	AF051347 Mus muscu
31	54	7.7	2269	10	AF097437	AF097437 Mus muscu
32	49.2	7.0	185306	2	AC013570	AC013570 Homo sapi
33	48.8	7.0	37224	3	AC007061	AC007061 Leishmani
34	48.2	6.9	36823	9	HSU209G1	Z68873 Human DNA s
35	46	6.6	154577	2	AL357046	AL357046 Homo sapi
36	46	6.6	154959	2	AC009282	AC009282 Homo sapi
37	46	6.6	185257	9	AL355305	AL355305 Human DNA
38	44.8	6.4	190544	9	AC005768	AC005768 Homo sapi
39	44.2	6.3	110000	2	LMFLCHR15_5	Continuation (6 of
40	44.2	6.3	172200	2	AC022626	AC022626 Homo sapi
41	43.2	6.2	9211	3	AB030033	AB030033 Dictyoste
42	43.2	6.2	210321	2	AC079438	AC079438 Mus muscu
43	42.8	6.1	1580	3	PFACSPH	J02695 Plasmodium
44	42.6	5.1	164025	2	AC025562	AC025562 Homo sapi
45	42.6	5.1	166922	2	AL512288	AL512288 Homo sapi

ALIGNMENTS

RESULT 1

AF187066

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

AF187066 Mus musculus p75NTR-associated cell death executor (Nade) mRNA, 11-JUN-2000

Complete cds.

AF187066

AF187066.1 GI:8452897

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 700)

Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,

Suvarito,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.

NADP, a p75NTR-associated cell death executor, is involved in

signal transduction mediated by the common neurotrophin receptor

p75NTR

J. Biol. Chem. 275 (23), 17566-17570 (2000)

20298829

```
REFERENCE 2 (bases 1 to 700)
AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
        Hanaoka,T., Li,Y., Irie,S. and Sato,T.
TITLE Direct Submission
JOURNAL Otolaryngology/Pathology, Columbia
UNIVERSITY Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
LOCATION/Qualifiers 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES
    source 1..700
    /organism="Mus musculus"
    /strain="BALB/c"
    /db_xref="taxon:10090"
    1..700
    /gene="Nade"
    177..551
    /gene="Nade"
    /note="Involved in the common neurotrophin receptor
    p75NTR-mediated signal transduction; NADE"
    /codon_start=1
    /product="p75NTR-associated cell death executor"
    /protein_id="AAF75131.1"
    /db_xref="GI:8452898"
    /translation="MANVHQENEEMEQPLNQEDRPVGGEGHQPAAANNNNHHNHN
    HNHRRGQARLAPNFRWAIPNRQMNDGLGGDDMEMEEMREIRKRLRELQRLNC
    LRILMGLSNHHHDEFCCLMP"
BASE COUNT 178 a 187 c 203 g 132 t
ORIGIN
Query Match 98.9%; Score 692; DB 10; Length 700;
Best Local Similarity 99.3%; Pred. No. 3.3e-171;
Matches 695; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 acgagcgtgtgccagcagctcgagagctcctctgtgcgcgcggtgtgagcgagcgccg 60
Db 1 ACAGAGCTGTGGCCAGCAGCTCGAGCTCTCTGTGCGCGCGCGGTGCGCAGCGGCGCG 60
QY 61 agcgagcgggacagattgactggaagcgagctgagctcagcgagcgagcggaattgacag 120
Db 61 AGCGAGCGGGACAGATTGACTTGAAGCGGAGAGTCCAGCGCGGCGGAATTCACAGG 120
QY 121 aggtactacgccgaagggataggccagaaatagcaaccaggaaacaaatctcatctg 180
Db 121 AGGACTTACGCCGACAGGATAGGCCAGAAATAGCAACACAGGAACAAAATCTCATCG 180
QY 181 ccaatgtccaccagaaacagagactgagctgagcagccctgagaaatgacaggaacacc 240
Db 181 CCAATGTCCACCAGGAAACGAAGAGATGGAGCAGCCCTGCAGAAATGGACAGGAAGAC 240
QY 241 gcctgtggaggaggtgagggccaccagcctgctgcaaacacaacaacaacaaccaca 300
Db 241 GCCCTGTGGGAGGAGGTGAGGGCCACCAGCTGTCTGCAAAACAACAACAACAACACCA 300
QY 301 accataaccacaaccaccaccagagcgagcgagctgcccgaacttcccttaacttccgat 360
Db 301 ACCATAAACCAACCAACCAACCAAGAGCGAGCTGCGCGACTTGCCTCCTACTTCCGAT 360
QY 361 gggccattcccaacaggcagatgagcggttgggtggagatggagatgatatggaaa 420
Db 361 GGGCCATTCCCAACAGGCGAGATGAATGAGCGGTGTGGTGGAGATGGAGATGATATGGAAA 420
QY 421 tgttcattgaggagatgagagagatcccgagaagacttaggagctagagctgagaaatt 480
Db 421 TGTTCATGGAGAGATGAGAGAGATCCGGAGAAAGCTTAGGGAGGCTACAGCTGAGAAAT 480
QY 481 gtctacgcatcttattggggggagctgtctaaaccacacagatcaccatgatgaattctgcc 540
Db 481 GTCTACGCATCTTATGGGGAGCTGTCTTAACCAACCAACCAACCAATCTTCC 540
QY 541 ttatgcttgacttgggtcattcccccttgagatccatctactgtactccgctgtagccc 600
Db 541 TTATGCTTGTGACTTGGGTCAATTTCCCCCTTGAGATCCACTGTGACTCCCGCTGTAGCCC 600
QY 601 ttctcctgcgcatcttctcctgacatgcctttaatgacccgcttgtgtgagccctgtgtat 660
us-09-327-750d-28.rge
Db 601 TTTTCTCGCAATTTCTTCTGACATGCTTTAATGACCCGTTTGTGGTGACCTTGTGTAT 660
QY 661 ttccatgccatgtgccaggtggtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 700
Db 661 TTCCATGCCATGTGCCAGGTGgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 700
RESULT 2
AF097440 854 bp mRNA ROD 13-APR-1999
LOCUS Mus musculus brain expressed X-linked protein 3 (Bex3) mRNA,
DEFINITION complete cds.
ACCESSION AF097440
VERSION AF097440.1 GI:4580593
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 854)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070
REMARK Erratum:([published erratum appears in Hum Mol Genet 1999
May;8(5):943])
REFERENCE 2 (bases 1 to 854)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
    source 1..854
    /organism="Mus musculus"
    /strain="FVB/N"
    /db_xref="taxon:10090"
    /db_xref="dbEST:AA272375"
    /chromosome="X"
    /map="near Plp"
    /tissue_type="pooled organs"
    1..854
    /gene="Bex3"
    172..546
    /gene="Bex3"
    /codon_start=1
    /product="brain expressed X-linked protein 3"
    /protein_id="AAD24431.1"
    /db_xref="GI:4580594"
    /translation="MANVHQENEEMEQPLNQEDRPVGGEGHQPAAANNNNHHNHN
    HNHRRGQARLAPNFRWAIPNRQMNDGLGGDDMEMEEMREIRKRLRELQRLNC
    LRILMGLSNHHHDEFCCLMP"
BASE COUNT 237 a 212 c 228 g 177 t
ORIGIN
Query Match 96.6%; Score 676; DB 10; Length 854;
Best Local Similarity 99.3%; Pred. No. 5.3e-167;
Matches 679; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 14 cagcagctcggagctcctctgtgcgcgcggtgtgagcgagcgagcgagcgagcgagcgag 73
Db 9 CAGCAGCTCGGAGCTCTCTGTGCGCGCGGCTGCCAGCGGCGCGCGCGCGGAG 68
QY 74 agattgactggaagccgaggtccagcgcgcggtggaattgacagagagactacgcgcg 133
Db 69 AGATTGACTGGAGCCGAGAGTCCAGCGCGGCGGGAATTGACAGAGGAGTACGCGCG 128
QY 134 aaggatagggccagaaatagcaaccaggaacaaatctcatcatgtgccaatgtccacca 193
Db 129 AAGGATAGGCCAGATAGCAACCAACCAAGGAACAAATCTCATCATGTGCCACCA 188
```



```
QY 194 ggaacacgaagctggagcagccctgcagaaatggacaggacacccgctgtgtggagg 253
Db 189 GGAACACGAAGAGATGAGCAGCCCTGCAGAAATGCACAGGAAGACCGCCCTGTGGAGG 248
QY 254 aggtgagggccacagcctgtgtgcaaaacaaacaaacacacacacacacacaa 313
Db 249 AGGTGAGGGCCACACAGCCTGTGCAAAACAAACAAACAAACAAACAAACAAACAA 308
QY 314 ccaccacgaagagccagcctgcgcagcctgcgccttaactccgatgggcatcccaa 373
Db 309 CCACCACCAAGAGCCAGGCTCGCGACTTGCCTTAATCCGATGGGCAATCCCA 368
QY 374 caggcagatgaatgacgggtgtgggtggagatggagatgatggaatgttcatggagga 433
Db 369 CAGGCAGATGAATGACGGGTTCGGTGGAGATGGAGATGATATGGAATTTTCATGGAGGA 428
QY 434 gatgagagatcccgagaaagcttagggagctacagctgagaaattgtctacgcatcct 493
Db 429 GATGAGAGAGATCCCGGAGAAAGCTTAGGGAGCTACAGCTGAGAAATTTGTTACGCATCCT 488
QY 494 tatggggagctgtctaaacacacacacacacacacacacacacacacacacacac 553
Db 489 TATGGGGAGCTGTAAACACACACACACACACACACACACACACACACACACACAC 548
QY 554 tcggttattcccccttgagatccatctgtgactcccgctgtagcccttcccttcgcat 613
Db 549 TCGGTATTCCTCCCTCCCTGAGATCCATCTGTGACTCCCGCTGTAGCCCTTTTCTCGCAT 608
QY 614 ttccgtgacatgctttaaataacccgtttgtgtgagccctgtgtatttccatgccat 673
Db 609 TTCCGTGACATGCTTTAATGACCCCTTTGTGTGAGCCTTGTATTATTCATGCCATGT 668
QY 674 gccagtgaggctgtgtgtggcag 697
Db 669 GCCAGTGGGGCTGTGTGGCCAG 692

RESULT 3
AF187065 519 bp mRNA 11-JUN-2000
LOCUS Rattus norvegicus p75NTR-associated cell death executor (Nade)
DEFINITION mRNA, complete cds.
ACCESSION AF187065
VERSION AF187065.1 GI:8452895
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 519)
AUTHORS Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,
Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.
TITLE Nade, a p75NTR-associated cell death executor, is involved in
signal transduction mediated by the common neurotrophin receptor
p75NTR
JOURNAL J. Biol. Chem. 275 (23), 17566-17570 (2000)
MEDLINE 20298829
REFERENCE 2 (bases 1 to 519)
AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
Hanaoka,T., Li,Y., Irie,S. and Sato,T.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES
source
1..519
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1..519
/gene="Nade"
44..406
/gene="Nade"
```

```
BASE COUNT 142 a 130 c 136 g 111 t
ORIGIN
Query Match 50.8%; Score 355.8; DB 10; Length 519;
Best Local Similarity 86.5%; Pred. No. 6.7e-83;
Matches 423; Conservative 0; Mismatches 47; Indels 19; Gaps 2;
QY 191 ccaggaacacgaagctggagcagccctgcagaaatggacaggacacccgctgtgtggagg 250
Db 28 CCAGGAAAACGAAGAAATGGAGCAGCCCTGCAGAAATGCACAGGAAGACCGCCCTGTGGG 87
QY 251 agaggtgagggccacacacacacacacacacacacacacacacacacacacacacac 310
Db 88 AGGAGGTGAGGGCCACACAGCCTGTGCAAAACAAACAAACAAACAAACAAACAAACAA 147
QY 311 caaccac-----caccgaagaggccagctgcgcagcctgcgcctta 352
Db 148 CAACCACACACACACACACCAATCATCACCGGAGAGGACAGGCCCGCCGACTTGCCTTAA 207
QY 353 ctccgatggggccattcccaacagcagatgaatgacgggttgggtggagatggagatga 412
Db 208 CTTTCGATGGGCATTCCTCCACAGGCAGATGAATGATGGTGTGGTGGAGATGAGATGA 267
QY 413 tatgaaatgttcatgagagatgagagatccgagagaaagcttagggagctacagct 472
Db 268 TATGGAATGTTTCATGGAGGAGATGAGAGATCCGGAGAAAGCTTAGGGAGCTGCAGTT 327
QY 473 gagaattgtctacgcatccttatggggagctgtcttaaccacacacacacacacacacac 532
Db 328 GAGAATGTCCTGCGTATTCCTTATGGGGAGGCTCTCTAATCACCAGCACCATCAGCATGA 387
QY 533 atttgcttattgcttgacttcggtcattccccctcgagatccatctgtgactcccg 592
Db 388 ATTCTGCTTATGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCT 446
QY 593 tctagcccttccctgcgcaatttctgcacatccctttaaagacccgttctgtgtgagccc 652
Db 447 TGTATTCCTGAGCCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCT 506
QY 653 tctgttatt 661
Db 507 CTTGGTGT 515

RESULT 4
BC003190 793 bp mRNA PRI 12-JUL-2001
LOCUS Homo sapiens, p75NTR-associated cell death executor: ovarian
DEFINITION granulosa cell protein (13Kd), clone MGC:802 IMAGE:3357965, mRNA,
complete cds.
ACCESSION BC003190
VERSION BC003190.1 GI:13112030
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 793)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
```

REMARK  
COMMENT

Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandohi, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 6 Row: k Column: 22  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA qi: 7657043.

FEATURES

```

Location/Qualifiers
1. 793
/organism="Homo sapiens"
/db_xref="LocusID:27018"
/db_xref="taxon:9606"
/clone="MGC:802 IMAGE:3357965"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH_MGC_16"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
196. 531

```

CDS

```

/codon_start=1
/product="p75NTR-associated cell death executor; ovarian
granulosa cell protein (13kD)"
/protein_id="AAH03190.1"
/db_xref="GI:1312031."
/translation="NANTHOENEEFPMQNGEDRPLGGCGEHQPNRGRGARLA
PNFRVALPNQINDMGCGDDMEIFMEEMREIRKRELQRLNCLRLMGCELSNIHHD
HDPFFCLMG"

```

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Query Match 48.3%; Score 337.8; DB 9; Length 793;  
Best Local Similarity 73.8%; Pred. No. 3.7e-78;

Matches	520;	Conservative	0;	Mismatches	132;	Indels	53;	Gaps	5;
QY	1	acagcgctgtgcccagcagctgcgagctctcttcgcgcgcggcgccgctgcgcaggcgggccccg	60						
Db	12	ACTCGGCTTGTGGTTACAGCTCCCGCTGCCCTTGAGCTCGCGGGCTTGGCATTCGGCCGC	71						
QY	61	aggcg-agcggggcacattgactggaagccgcagaagtccaggcgca-----gcggaaa	112						
Db	72	GGAAAGAAGCGCAGAGCTTCGCAGGCTAAGTGTCTCCGCGGCGCACCTCGCGCGAGAA	131						
QY	113	ttagcaggaggactagccgcagagggatagcccgagaataagcaaccaggaaacaatact	172						
Db	132	TCCGGAGGAGGAAGGAGACTGCAAGGATAGGCCCAGAAAAACAACCAGAAAAAAAATCT	191						
QY	173	catcatggccaatgtgccaccaggaaacaagaagcttgagcagccctcgagaatggaca	232						
Db	192	CATCATTGGCAAAATTTCACAGAAAACGAAGAGATGGAGCAGCCTATGCGAGAAATGGAGA	251						
QY	233	ggaacaccgcacctgtgggagaggtgagggccaccagcctgtctgcaaacacaacaaca	292						
Db	252	GGAAGACCGCCCTTTGGGAGGAGGTGNAGGCCACCAACCTTGCAGGAA-----	298						
QY	293	caaccacaacca taacacaacaccacacogagagcgaggtctgcgcagctgcccttaa	352						

Db	299	-----ATCGACGGGGACAGGCTCGCGACTTGCCCCCTAA	332
Qy	353	cttcgagggccattcccaacagcgagatgacgaggggtgggtgagatgagatga	412
Db	333	TTTTTCGATGGCCATACCCAATAGCGAGATCAATGATGGGATGGGTGGAGATGAGATGA	392
Qy	413	tatgaaatgttcagtgagagagatgagagatcccgagaaagcttaggagctacagct	472
Db	393	TATGAAATATTATCGAGGAGATGAGAGAAATCAGAGAAACTTATAGGAGCTTGCAGTT	452
Qy	473	gagaaatgtctacgacatccttatgggggagctgtctaacaccacacgatacaccatgata	532
Db	453	GAGAAATGTGTCGCTATGCTATATGGGGAGCTCTCTAATCACCATGACCATCATGATGA	512
Qy	533	attgccttatgcttgactcgggtcatctccccctcgagatccatcactatgactccgc	592
Db	513	ATTTTGGCTTATGCTTGACTCTGCCATT-TATCATGAGATTAAATGATGATTCGCCGC	571
Qy	593	tgtagccttccctcgcatttcttcctgacatgcctttaatgacccgcttggtaggcgcc	652
Db	572	TGTTTCTTTTCTTCCTGCATTTCCTAATATGCTTACTGATCCGTTTCTGCTGGAACCC	631
Qy	653	tgtgtatttcacatgcacatgcccaggtgccaggtggggctgtgctgtgcag	697
Db	632	TATGTTATT-CCATGTGTCAAGTGGGTCTTGTGTGCCAG	671
RESULT	5		
HUMOGC			
LOCUS		891 bp mRNA	PRI 07-MAR-1995
DEFINITION		Human unknown protein from clone pHGR74	mRNA, complete cds.
ACCESSION		M38188 X56942	
VERSION		M38188.1	GI:189378
KEYWORDS			
SOURCE		Human ovarian granulosa cell line, cDNA to mRNA.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		Rapp,G., Freudenstein,J., Klaudiny,J., Mucha,J., Wempe,F., Zimmer,M. and Scheit,K.H.	
JOURNAL		DNA Cell Biol. 9 (7), 479-485 (1990)	
MEDLINE		Characterization of three abundant mRNAs from human ovarian granulosa cells	
COMMENT		Draft entry and computer-readable sequence for IDNA 9, 479-485 (1990) kindly submitted by K.H.Scheit, 27-AUG-1990.	
FEATURES		Location/Qualifiers	
source		1..891	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="pHGR74"	
		/cell_type="granulosa"	
		/tissue_type="ovary"	
mRNA		<1..891	
		/note="protein of unknown function"	
CDS		312..647	
		/note="protein of unknown function"	
		/codon_start=1	
		/protein_id="AA663232.1"	
		/db_xref="GI:189379"	
		/translation="MANIQHEMEMEOPMQNEEDRPLCGGGGHPAGNRRGQARRLA	
		PNFRAIPNRQINDMGGGDDWEIFMEEMREIRKRLRELQRLNLRILMGLSNHHD	
		HHDEFCLMP	
		361..534	
		/note="protein of unknown function"	
CDS		/codon_start=1	
		/protein_id="AA663233.1"	
		/db_xref="GI:189380"	
		/translation="MERKTALEEVKATSLQIEDGRDLADLPIDGPPYPIGRSMGVM	
		VEEMWIKYSWRR	
BASE COUNT	251 a	182 c	234 t

ORIGIN	FEATURES	Source
Query Match Best Local Similarity 43.9%; Score 307.4; DB 9; Length 891; Matches 428; Conservative 0; Mismatches 76; Indels 45; Gaps 3;	Location/Qualifiers 1..891 /organism="Homo sapiens" /db_xref="taxon:9606" 1..891 /gene="NADE" 312..647 /gene="NADE" /note="involved in the common neurotrophin receptor p75NTR-mediated signal transduction" /codon_start=1 /product="p75NTR-associated cell death executor" /protein_id="AAF75129.1" /db_xref="GI:8452894" /translation="MANIHOENEEMEQMONGEEDRPLGGGEGHOPAGNRRGQARRLA PNFRWAIPINRQINDMGDDMEIFMEEMREIRKLRELQLNRLILMGLSNHHD HHDEFCLMP"	University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
QY 149 aaatagcaaccaggaaacaaatctcatcgccaatgtccaccaggaaacaaagagct 208       Db 284 AAAACAACACAGAAAAAATCTCATGCGCAATATTCACCGAGAAACGAGAGAT 343	BASE COUNT 251 a 182 c 224 g 234 t ORIGIN	
QY 209 gggagcccccctgcagaaatggacagaaacacccctctgtggaggaggtgagggccacca 268       Db 344 GGAGCAGCCTATGCAGAAATGGAGAGAGACCCCTTTGGGAGGAGGTGAAGCCACCA 403	Query Match 43.9%; Score 307.4; DB 9; Length 891; Best Local Similarity 78.0%; Pred. No. 3.6e-70; Matches 428; Conservative 0; Mismatches 76; Indels 45; Gaps 3;	
QY 269 gctctgctgaacaaacaaacaaacacacacacacacacacacacacacacacacagag 328       Db 404 GCTGCAGGAA-----ATCGACGGGG 424		
QY 329 ccaggctcggacgtccccccttaacttcgagtgccattcccaacaggcagatgaatga 388       Db 425 ACAGGCTCGCGACTTGCCTTAATTTTCGATGGGCCATACCAATAGGCAGATCAATGA 484		
QY 389 cgggttgggtggagatggagatgataatgttcacgagagatgagagagatgccg 448       Db 485 TGGGATGGGTGGAGATGGAGATGATGGAATATTCATGAGGAGATGAGAGAAATCAG 544		
QY 449 gagaaagcttagggagagctacagctcagagaaatgtctacgacatccttatggggagctg 508       Db 545 AAGAAACTTAGGGAGCTGAGGAGTTCAGTTCGCGTATCTCTGCGTATCTTATGGGGAGCTCTC 604		
QY 509 taaccaccagatcacatgatgaattctgccttatgccttgacttcggtcattcccccc 568       Db 605 TAATCACCATGACCATCATGATGAATTTGCGCTTATGCGCTTGACTCCTGCCATT-TATCA 663		
QY 569 tgagatcccatctgtgactccccctgtagcctttcccttcgcattttcctgcacatgcc 628       Db 664 TGAGATTAATACTGTGATTCGCCCTGTTCTTTTTCCTTGCCATTTCCTAATATGCCIT 723		
QY 629 taatgaccggttgggtggagccctgtgtatttccatgccatgctgcccaggtgggcttg 688       Db 724 TACTGATCGGTTTGGTGTGAACCTATGTTATTT-----CCATGTGTCAAGTGGGTCTTG 778		
QY 689 tgttgccag 697       Db 779 TGTGCCAG 787		
RESULT 6 AF187064 891 bp mRNA PRI 11-JUN-2000 LOCUS Homo sapiens p75NTR-associated cell death executor (NADE) mRNA, complete cds. ACCESSION AF187064 VERSION AF187064.1 GI:8452893 KEYWORDS SOURCE human. ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. REFERENCE 1 (bases 1 to 891) AUTHORS Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D., Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A. NADE, a p75NTR-associated cell death executor, is involved in signal transduction mediated by the common neurotrophin receptor p75NTR J. Biol. Chem. 275 (23), 17566-17570 (2000) MEDLINE 20298829 REFERENCE 2 (bases 1 to 891) AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P., Hanaoka,T., Li,Y., Irie,S. and Sato,T. Direct Submission TITLE Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia		
QY 149 aaatagcaaccaggaaacaaatctcatcgccaatgtccaccaggaaacaaagagct 208       Db 284 AAAACAACACAGAAAAAATCTCATGCGCAATATTCACCGAGAAACGAGAGAT 343		
QY 209 gggagcccccctgcagaaatggacagaaacacccctctgtggaggaggtgagggccacca 268       Db 344 GGAGCAGCCTATGCAGAAATGGAGAGAGACCCCTTTGGGAGGAGGTGAAGCCACCA 403		
QY 269 gctctgctgaacaaacaaacaaacacacacacacacacacacacacacacacagag 328       Db 404 GCTGCAGGAA-----ATCGACGGGG 424		
QY 329 ccaggctcggacgtccccccttaacttcgagtgccattcccaacaggcagatgaatga 388       Db 425 ACAGGCTCGCGACTTGCCTTAATTTTCGATGGGCCATACCAATAGGCAGATCAATGA 484		
QY 389 cgggttgggtggagatggagatgataatgttcacgagagatgagagagatgccg 448       Db 485 TGGGATGGGTGGAGATGGAGATGATGGAATATTCATGAGGAGATGAGAGAAATCAG 544		
QY 449 gagaaagcttagggagagctacagctcagagaaatgtctacgacatccttatggggagctg 508       Db 545 AAGAAACTTAGGGAGCTGAGGAGTTCAGTTCGCGTATCTCTGCGTATCTTATGGGGAGCTCTC 604		
QY 509 taaccaccagatcacatgatgaattctgccttatgccttgacttcggtcattcccccc 568       Db 605 TAATCACCATGACCATCATGATGAATTTGCGCTTATGCGCTTGACTCCTGCCATT-TATCA 663		
QY 569 tgagatcccatctgtgactccccctgtagcctttcccttcgcattttcctgcacatgcc 628       Db 664 TGAGATTAATACTGTGATTCGCCCTGTTCTTTTTCCTTGCCATTTCCTAATATGCCIT 723		
QY 629 taatgaccggttgggtggagccctgtgtatttccatgccatgctgcccaggtgggcttg 688       Db 724 TACTGATCGGTTTGGTGTGAACCTATGTTATTT-----CCATGTGTCAAGTGGGTCTTG 778		
QY 689 tgttgccag 697       Db 779 TGTGCCAG 787		
RESULT 6 AF187064 891 bp mRNA PRI 11-JUN-2000 LOCUS Homo sapiens p75NTR-associated cell death executor (NADE) mRNA, complete cds. ACCESSION AF187064 VERSION AF187064.1 GI:8452893 KEYWORDS SOURCE human. ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. REFERENCE 1 (bases 1 to 891) AUTHORS Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D., Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A. NADE, a p75NTR-associated cell death executor, is involved in signal transduction mediated by the common neurotrophin receptor p75NTR J. Biol. Chem. 275 (23), 17566-17570 (2000) MEDLINE 20298829 REFERENCE 2 (bases 1 to 891) AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P., Hanaoka,T., Li,Y., Irie,S. and Sato,T. Direct Submission TITLE Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia		

RESULT 7  
LOCUS G72708/c  
DEFINITION G72708 421 bp DNA STS 08-AUG-2001  
MARC 4953-4954:991939031:1 SCF - porcine spleen Sus scrofa Sfs  
genomic, sequence tagged site.  
ACCESSION G72708





```
misc_feature /note="MER58A repeat: matches 9. .214 of consensus"
13338. .13833 /note="match: GSS: Em:AQ568864"
repeat_region 13688. .13749
repeat_region /note="MIR repeat: matches 77. .139 of consensus"
13840. .13966 /note="MIR repeat: matches 22. .147 of consensus"
repeat_region 14162. .14245
repeat_region /note="MIR repeat: matches 46. .136 of consensus"
14925. .15058 /note="MIR repeat: matches 48. .191 of consensus"
repeat_region 15994. .16040 /note="MIR repeat: matches 102. .144 of consensus"
repeat_region 16116. .16184
repeat_region /note="MER58B repeat: matches 3. .70 of consensus"
16897. .17005 /note="MIR repeat: matches 9. .118 of consensus"
repeat_region 17618. .17700
repeat_region /note="MIR repeat: matches 108. .192 of consensus"
17748. .17912
repeat_region /note="MIR repeat: matches 73. .245 of consensus"
18657. .18871 /note="MIR repeat: matches 35. .252 of consensus"
repeat_region 18927. .19230
repeat_region /note="AluX repeat: matches 1. .301 of consensus"
19492. .19559 /note="MIR repeat: matches 2. .67 of consensus"
repeat_region 19560. .19821
repeat_region /note="AluY repeat: matches 34. .295 of consensus"
19822. .19959
repeat_region /note="MIR repeat: matches 67. .212 of consensus"
19987. .20286
repeat_region /note="AluY repeat: matches 1. .300 of consensus"
20316. .20512
repeat_region /note="MIR repeat: matches 82. .262 of consensus"
20513. .20741
repeat_region /note="L1MD1 repeat: matches 5970. .6224 of consensus"
20769. .21074
repeat_region /note="AluSg1 repeat: matches 1. .303 of consensus"
21085. .21207
repeat_region /note="L1M2 repeat: matches 1776. .1898 of consensus"
21212. .21391
repeat_region /note="L1MD repeat: matches -6. .178 of consensus"
21392. .21471
repeat_region /note="MIR repeat: matches 3. .86 of consensus"
22146. .22441
repeat_region /note="AluX repeat: matches 1. .296 of consensus"
22526. .22610
repeat_region /note="LTR16C repeat: matches 267. .349 of consensus"
22940. .23204
repeat_region /note="HERV16 repeat: matches 378. .650 of consensus"
23205. .23514
repeat_region /note="AluSg repeat: matches 1. .311 of consensus"
23919. .24194
repeat_region /note="AluX repeat: matches 36. .311 of consensus"
24496. .24590
repeat_region /note="LTR16C repeat: matches 253. .349 of consensus"
24871. .24916
repeat_region /note="MIR repeat: matches 102. .143 of consensus"
25378. .25555
repeat_region /note="L2 repeat: matches 2330. .2519 of consensus"
25673. .25910
repeat_region /note="MIR repeat: matches 4. .255 of consensus"
26722. .26798
repeat_region /note="L2 repeat: matches 2572. .2649 of consensus"
27040. .27107
repeat_region /note="L2 repeat: matches 2489. .2500 of consensus"
27108. .27374
repeat_region /note="AluY repeat: matches 41. .304 of consensus"
27375. .27471
repeat_region /note="L2 repeat: matches 2372. .2489 of consensus"
27651. .27945
repeat_region /note="AluX repeat: matches 1. .299 of consensus"
```

```
repeat_region 28235. .28876 /note="L2 repeat: matches 1113. .1817 of consensus"
repeat_region 28975. .29097 /note="MER96 repeat: matches 42. .173 of consensus"
repeat_region 29173. .29196 /note="L2 copies 2 mer aa 100 conserved"
misc_feature 29832. .30260 /note="match: GSS: Em:AQ075670"
repeat_region 30392. .30597 /note="MIR repeat: matches 35. .252 of consensus"
repeat_region 30845. .31010 /note="MER91B repeat: matches 2. .162 of consensus"
repeat_region 31056. .31094 /note="MIR repeat: matches 218. .256 of consensus"
repeat_region 31728. .31791 /note="L6 copies 4 mer t9tg 76 conserved"
repeat_region 31858. .31952 /note="MIR repeat: matches 48. .153 of consensus"
misc_feature complement(33437. .33589) /note="match: STS: Em:H55731"
repeat_region 33543. .33633 /note="MIR repeat: matches 84. .165 of consensus"
repeat_region 33785. .33966 /note="MER91A repeat: matches 1. .185 of consensus"

Query Match 22.1%; Score 154.6; DB 9; Length 98274;
Best Local Similarity 76.5%; Pred. No. 7e-30; Mismatches 0; Indels 14; Gaps 6;
Matches 270; Conservative 0;

QY 345 gccctaaattccgagtgccattcccaacaggcagatgaatgacgggttggtggagat 404
DB 62890 GCCCCTAATTTCAATGGCCATACCCATAGCAGGTCATGGATGAATAGAGGT 62949

QY 405 gaagatgatgaaatgttcattgagagatgagagagatccgcgaagaagcttagggag 464
DB 62950 GGAGATGATATGCAATGTTTCATGGAGGAGATGAGAGGAATCAGGAG--AGAATAAGGAG 63007

QY 465 ctacagctgagaaattgtctacgcatcttggggagctgtctaaacaccagctac 524
DB 63008 CTACAAATGAGGAATGTCTGTATCCTTAT--GGGAAGCTGTGTAATCCCCATGACCAT 63066

QY 525 catgatgaattgccttatgccttgacttcggttcattcccccctgagatccatactgtg 584
DB 63067 CATGATGAATTTGCC--AATGCTGACTCTCCATTT--TTCCATGAGGTTAATATTGTG 63124

QY 585 atccgcctgtgaccttccctgcatttccctgcacatccttgaacgccttgaacgcctgttg 644
DB 63125 ATTCCCATTTGTTTCTTTTTCCTTCATTTCTCAATATGCTCTTT---ACTCATTTGCT 63180

QY 645 gtgagccctgttatttccatgccatgtgccagatggggctgtgtgtgtgcccag 697
DB 63181 GTGAACCTTATGTTATTT-----CCATGTGTCAGGTGAGTCTTTGTCTTCCAG 63228

RESULT 10
G35294/c 477 bp DNA STS 02-OCT-1997
LOCUS human STS SHGC-37409, sequence tagged site.
DEFINITION G35294
ACCESSION G35294
VERSION G35294.1 GI:2459462
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS Myers,R.M.
JOURNAL Unpublished (1997)
COMMENT Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
```







```
Db 256 AGGAGGGGAGCAGCCCCACGACAGATGAAGAAATCCCGCCATTTGGGAGGGGTGA 315
Qy 260 gggccaccagcctgctgcaaaacaaacaaacacacacacacacacacacacacac 319
Db 316 AGCCAGAGGCTGGAGAAATATCA----- 341
Qy 320 ccgaagagccaggctgcccacctgcccctaaacttcgatggccattcccaacagga 379
Db 342 --GGCGGGGGAGTTAGCGGACTTGTCCCTAATTTTCGATGGCGCATACCTTAATAGGCA 399
Qy 380 gatgaatgacgggttgggtggagatggagatgataatggaattgttcacatgagatgag 439
Db 400 TATTGAGCACAA-----TGAAGCGAGAGATGATGAGAAAGTTTGGTGGCGCATATGAT 453
Qy 440 agagatcgcggagaagacttaggagctacagctgagaaattgtctacacatccttat99g 499
Db 454 GGAATCAAGAAAGACTAGGGAACAGCAGATGAGGCACATATATGCGCTTCCAAA---- 509
Qy 500 ggagctgtctaaccaccagcatcaccatgactccgcctgtagccctttccctcgcatttcc 559
Db 510 -----CTCCTGAACCTGACAACCATATGACTTTTGCCTCATACCTTGAATCCTAAA 561
Qy 560 attccccctgagatccatactgtgactccgcctgtagccctttccctcgcatttcc 618
Db 562 AGTTTCGCTGAGGTTAATGTGAACACTGCTTTTACAAGCTTGTATTTTGTGATTACT 620

RESULT 13
AX100231
LOCUS AX100231 1364 bp DNA PAT 02-APR-2001
DEFINITION Sequence 32 from Patent WO0119860.
ACCESSION AX100231
VERSION AX100231.1 GI:13539115
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Tang,Y.T., Hillman,J.L., Yue,H., Reddy,R., Lal,P., Shah,P.,
Aizimzai,Y., Baughn,M.R., Lu,D.A., Bandman,O., Shih,L.L. and
Patterson,C.
TITLE Proteins associated with cell differentiation
JOURNAL Patent: WO 0119860-A 32 22-MAR-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source 1..1364
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 1990956CBI"
BASE COUNT 411 a 269 c 322 g 362 t
ORIGIN

Query Match 11.5%; Score 80.2; DB 6; Length 1364;
Best Local Similarity 56.2%; Pred. No. 1.8e-10;
Matches 269; Conservative 0; Mismatches 153; Indels 57; Gaps 4;

Qy 143 gccagaatagcaaacaggaaacaaatctcatcgtgccaatgtctccaccaggaacagca 202
Db 255 GTCCAAAGAGGAACATAGCGGCAAAACAATCTCAACGGGGAATAATGCCCAACAGAAACGA 314
Qy 203 agagctggagcag---ccctcgagaatggacagaaacccgcctgtgggagagggtga 259
Db 315 AGAGGGGAGCAGGCCCCCACCACAGAAAGAAATCCGCCCATTTGGGAGGGGTGA 374
Qy 260 gggccaccagcgtgtgtcaaaacaaacaaacacacacacacacacacacacacac 319
Db 375 AGCCAGAGGCTGGAGGAATATCA----- 400
Qy 320 ccgaagagccaggctgcgcacctgcccctaaacttcgatggccattcccaacagga 379
Db 401 --GGCGGGGGAGTTAGCGGACTTGTCCCTAATTTTCGATGGCGCATACCTTAATAGGCA 458
```

```
Qy 380 gatgaatgacgggttgggtggagatggagatgatggaattgttcacatgagagatgag 439
Db 459 TATTGAGCACAA-----TGAAGCGAGAGATGATGAGAAAGTTTGTAGGCGCATGAT 512
Qy 440 agagatcgcggagaagcttaggagctacagctagaaattgtctacacatccttat99g 499
Db 513 GGAATCAAGAAAGACTAGGGAACAGCAGATGAGGCACATATATGCGCTTCCAAA---- 568
Qy 500 ggagctgtctaaccaccagcatcaccatgactccgcctgtagccctttccctcgcatttcc 559
Db 569 -----CTCCTGAACCTGACAACCATATGACTTTTGCCTCATACCTTGAATCCTAAA 620
Qy 560 attccccctgagatccatactgtgactccgcctgtagccctttccctcgcatttcc 618
Db 621 AGTTTCGCTGAGGTTAATGTGAACACTGCTTTTACAAGCTTGTATTTTGTGATTACT 679

RESULT 14
HS635G19
LOCUS HS635G19 69648 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 635G19 on chromosome Xq22.1-22.3
Contains a LAMR1 (Laminin Receptor 1 (67kD) (RPSA, 40S Ribosomal
Protein SA, P40)) pseudogene and part of a novel protein. Contains
ESTs and GSSs, complete sequence.
ACCESSION AL035494
VERSION AL035494.8 GI:4775622
KEYWORDS HTG; 40S Ribosomal Protein P40; 40S Ribosomal Protein SA; CpG
Island; Laminin Receptor 1; LAMR1; RPSA.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 69648)
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On May 11, 1999 this sequence version replaced gi:4678749.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
635G19 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone 635G19.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone U65A4 (Z81014) is at 69549 in this
sequence. The true right end of clone U101D3 (Z85997) is at 100 in
this sequence.
Location/Qualifiers
1..69648
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```
/chromosome="X"
/map="q22.1-22.3"
/clone="RP4-635G19"
/clone_lib="RPCI-4"
439..483
/notes="TAR1 repeat: matches 707..749 of consensus"
501..661
/notes="AluJb repeat: matches 137..311 of consensus"
662..897
/notes="118 copies 2 mer aa 70% conserved"
915..1331
/notes="MSTA repeat: matches 1..419 of consensus"
1334..1517
/notes="92 copies 2 mer ga 77% conserved"
1927..2020
/notes="MIR repeat: matches 109..217 of consensus"
2066..2262
/notes="LIME2 repeat: matches 5825..6022 of consensus"
2263..2574
/notes="AluSp repeat: matches 1..313 of consensus"
2575..2819
/notes="LIME2 repeat: matches 5583..5825 of consensus"
2821..3039
/notes="Alu repeat: matches 1..311 of consensus"
3104..3384
/notes="LIME2 repeat: matches 5304..5591 of consensus"
3391..4007
/notes="L1MB1 repeat: matches 5558..6162 of consensus"
4012..4137
/notes="trigger2a repeat: matches 11..123 of consensus"
4141..4252
/notes="L1 repeat: matches 2920..3031 of consensus"
4265..4350
/notes="HY4 repeat: matches 1..87 of consensus"
4350..4678
/notes="MER74A repeat: matches 28..369 of consensus"
4914..5095
/notes="MER5A repeat: matches 8..189 of consensus"
5224..5555
/notes="L2 repeat: matches 2423..2746 of consensus"
6040..6086
/notes="L2 repeat: matches 1676..1720 of consensus"
7413..7478
/notes="Alu repeat: matches 239..304 of consensus"
7628..7928
/notes="AluYb8 repeat: matches 1..316 of consensus"
7952..7987
/notes="18 copies 2 mer aa 81% conserved"
7990..8305
/notes="AluJo repeat: matches 1..309 of consensus"
8436..8538
/notes="MIR repeat: matches 99..212 of consensus"
10836..11442
/notes="L1MD repeat: matches 1357..1972 of consensus"
11443..11571
/notes="FLAM_A repeat: matches 2..129 of consensus"
11572..11946
/notes="L1MD repeat: matches 977..1357 of consensus"
11947..12245
/notes="AluY repeat: matches 1..301 of consensus"
12246..13098
/notes="L1MD repeat: matches 275..977 of consensus"
13094..13223
/notes="L1MD repeat: matches -17..113 of consensus"
13523..13820
/notes="MLTID repeat: matches 177..505 of consensus"
13823..13932
/notes="55 copies 2 mer ct 77% conserved"
13936..13977
/notes="MLTID repeat: matches 201..241 of consensus"
13978..14240
/notes="AluSg1 repeat: matches 1..270 of consensus"
14245..14534
/notes="AluSc repeat: matches 1..290 of consensus"
14535..14721
/notes="MLTID repeat: matches 2..201 of consensus"
14724..15604
/notes="L1MC/D repeat: matches 4646..5565 of consensus"
15605..15909
/notes="AluX repeat: matches 24..312 of consensus"
16069..16363
/notes="AluX repeat: matches 1..291 of consensus"
16482..16641
/notes="AluJb repeat: matches 136..295 of consensus"
16671..16953
/notes="AluX repeat: matches 1..284 of consensus"
17164..17280
/notes="L1P repeat: matches 4499..4619 of consensus"
17326..17377
/notes="26 copies 2 mer at 79% conserved"
17704..17863
/notes="AluSc repeat: matches 118..295 of consensus"
17864..18175
/notes="AluYb8 repeat: matches 1..318 of consensus"
18176..18298
/notes="AluSc repeat: matches 1..118 of consensus"
1848..18873
/notes="13 copies 2 mer tg 100% conserved"
18876..19171
/notes="AluSg repeat: matches 1..295 of consensus"
19183..19232
/notes="25 copies 2 mer ac 100% conserved"
19246..19533
/notes="AluY repeat: matches 5..292 of consensus"
19539..19837
/notes="MER74B repeat: matches 61..368 of consensus"
20016..20194
/notes="MER74A repeat: matches 295..487 of consensus"
20394..20692
/notes="AluSg repeat: matches 1..302 of consensus"
20701..20809
/notes="LTR1 repeat: matches 682..785 of consensus"
20877..21061
/notes="LTR1 repeat: matches 1..195 of consensus"
21186..21636
/notes="LTR29 repeat: matches 3..498 of consensus"
21609..21737
/notes="MER34 repeat: matches 409..543 of consensus"
21771..22076
/notes="AluY repeat: matches 2..308 of consensus"
22878..22974
/notes="MIR repeat: matches 92..191 of consensus"
23452..23582
/notes="FLAM_C repeat: matches 1..131 of consensus"
23583..23624
/notes="21 copies 2 mer ta 76% conserved"
23735..24041
/notes="AluSg1 repeat: matches 1..308 of consensus"
24045..24072
/notes="14 copies 2 mer tt 89% conserved"
25274..25582
/notes="AluJb repeat: matches 1..289 of consensus"
25583..25747
/notes="match: GSS B63622"
25959..26012
/notes="HERVPH21 repeat: matches 23..79 of consensus"
26144..26456
/notes="AluY repeat: matches 1..311 of consensus"
27281..27432
/notes="AluJb repeat: matches 137..289 of consensus"
27436..27507
/notes="L1M4 repeat: matches 5360..5434 of consensus"
27536..27600
/notes="L1M4 repeat: matches 2220..2283 of consensus"
27601..27975
/notes="MLT1A repeat: matches 3..365 of consensus"
```

repeat\_region 27976. .28273

```
Query Match 11.5%; Score 80.2; DB 9; Length 69648;
Best Local Similarity 56.2%; Pred No. 2.4e-10;
Matches 269; Conservative 0; Mismatches 153; Indels 57; Gaps 4;

QY 143 gccagaatagcaaccaggaaacaaatctcatcgtgccaatgtccaccaggaaacga 202
Db 64298 GTCCAAAGAGGAAC TAGCGGCAACAATCTCAACGGGGAATAATGCCCAACAAGAAACGA 64357

QY 203 agagctggagcag---ccctgcagaatggacagaaacacccctgtgtggaggaggga 259
Db 64358 AGGAGGGAGCAGGCGCCACACGACGAATGAAGAAGATCCCGCCATTTGGGAGGGGTGA 64417

QY 260 gggccaccagcctgctgcaacaaacaaacaaacacacacacacacacacacac 319
Db 64418 AGCCAGAGCCTGGAGGAATATCA----- 64443

QY 320 ccgaagagccaggtctgcgcacctgcccttaactccgatggccattcccaacaggca 379
Db 64444 --GGCGGGGGCGAGTTAGGCGACTTGTCCCTAATTTTCGATGGCCATACCTAATAGGCA 64501

QY 380 gatgaatgacgggttgggtggagatggagatgatgaaatgttcattgagagagatgag 439
Db 64502 TATTGAGCACA-----TGAAGCGAGAGATGATGTAGAAAGGTTTGTAGGCGAGATGAT 64555

QY 440 agagatccggagaaagcttaggagctacagctgagaaattgtctacgcacatcctatggg 499
Db 64556 GGAATCAAGAGAAGACTAGGGAACAGCAGATGAGGCACATATATGCGGTTCCAAA---- 64611

QY 500 ggaagtgttaaccaccagcatcaccatgatgaattctgccttatgccttgccttcggctc 559
Db 64612 -----CTCCTGAACCTGACAACCATATATGACATTTTCCCTCATACCTTGAATCCTAAA 64663

QY 560 attccccctgagatccatactgtgactcccgctgtagcccttccctgcgactttcct 618
Db 64664 AGTTTTCGTAGGTTAATGTGAACACTGCTTTTACAAGCTTGTATTTTGTGATTACT 64722

RESULT 15
LOCUS G13550 244 bp DNA STS 04-JAN-1996
DEFINITION human STS SHGC-11140.
ACCESSION G13550
VERSION G13550.1 GI:1129289
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; mitochondria; eukaryotes; Metazoa; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Carnivora; Hominidae; Homo.
1 (bases 1 to 244)
Myers.R.M.
Unpublished (1995)

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 415/259687
Fax: 415/259689
Email: myers@shgc.stanford.edu

Primer A: CCGTTTCCTGTGAACCCCTAT
Primer B: TTTAAATTTTTCATGCAATGG
STS size: 138
PCR Profile:
Initial incubation: 94 degrees C for 90 seconds
Denaturation: 94 degrees C for 15 seconds
Annealing: 62 degrees C for 23 seconds
Polymerization: 72 degrees C for 30 seconds
```

```
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600
Protocol:
Template: 25 ng each 1 uM
Primers: each 200 uM
Taq Polymerase: 0.05 units/ul
Total Vol: 10 ul

Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.3

Prepared with primer pairs derived from M38188 -- Unigene.
FEATURES
source
Location/Qualifiers
1..244
/organism="Homo sapiens"
STS
primer_bind 84..221
primer_bind 84..103
primer_bind complement(199..221)
BASE COUNT 57 a 50 c 42 g 95 t
ORIGIN

Query Match 11.1%; Score 77.4; DB 11; Length 244;
Best Local Similarity 80.2%; Pred. No. 8.6e-10;
Matches 105; Conservative 0; Mismatches 21; Indels 5; Gaps 1;

QY 567 cctgagatccatactgtgactcccgctgtagcccttccctgcgactttcctgcacatgcc 626
Db 15 CATGAGATTAATAC TGTGATCCCGCTGTTTCTTTTCTTCCATTTCCTAATATGCC 74

QY 627 ttaatgacccgtttgtgtgagccctgtgtatttccatgccaatgccaagtgagggtc 686
Db 75 TTTACTGATCCGTTTGTCTGTAACCCCTATGTTATTT-----CCATGTGTCAAGTGGGTCT 129

QY 687 tgtgttgcag 697
Db 130 TGTGTGCGCAG 140

Search completed: March 11, 2002, 16:33:03
Job time: 8158 sec
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2002, 13:08:25 ; Search time 1609.39 Seconds  
(without alignments)  
4673.848 Million cell updates/sec

Title: US-09-327-750D-28  
Perfect score: 700  
Sequence: 1 acagagctgtgccagcagc.....ggggctgtgtgccagtga 700

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	687.4	98.2	809	12	AK004531		Mus muscu
2	687.4	98.2	845	12	AK010500		Mus muscu
3	637	91.0	787	11	BF608455		MVL_00139
4	627.6	89.7	642	10	AI118980		ue93c08.y
5	606.8	86.7	628	10	AW476468		uq76d02.y
6	591	84.4	599	11	BG088461		H3153D07-
7	590.4	84.3	612	10	AA272375		vb62g07.r
8	585.4	83.6	826	12	AK003294		Mus muscu
9	578	82.6	817	11	BI407164		602919012
10	571.6	81.7	585	10	BE334877		us90b10.y
11	566.6	80.9	923	11	BF178306		601808545
12	535.6	76.5	542	10	BE334866		us90a11.y

13	533.2	76.2	810	11	BF168928		601775392
14	533.2	76.2	918	11	BI249717		602996190
15	531.4	75.9	557	10	AA215070		mu76f04.r
16	524.8	75.0	577	10	AA268306		va89c10.r
17	523.6	74.8	538	10	AI008575		ue15e06.y
18	516	73.7	590	11	W64711		md71g01.r1
19	510.4	72.9	547	10	AA476461		uq76c02.y
20	509.8	72.8	543	10	AA050176		mj14b06.r
21	509	72.7	577	11	BF681983		602116954
22	507	72.4	536	11	W98871		mf89d04.r1
23	507	72.4	616	11	W81757		me95d06.r1
24	497.4	71.1	542	10	AA004191		mg82d06.r
25	496.8	71.0	1042	11	W15845		mb53b09.r1
26	492.4	70.3	503	10	AW908751		uf57a05.y
27	489.2	69.9	853	11	BI082462		602877692
28	487	69.6	522	10	AA068855		mm64d09.r
29	485.4	69.3	506	10	AA049613		mj36b06.r
30	485.4	69.3	603	10	AW557617		L0284B06-
31	484	69.1	492	11	W46041		mc82h02.r1
32	483.6	69.1	533	10	AA606578		vm82d02.r
33	480	68.6	532	10	AA214909		mu76f05.r
34	479.8	68.5	534	10	AA259353		va50g11.r
35	476.6	68.1	533	11	W29675		mc07d02.r1
36	468.4	66.9	489	10	AA044548		mj12c02.r
37	466.6	66.7	560	10	AA870604		vq23g08.r
38	465	66.4	486	10	AA259786		va87c12.r
39	461.2	65.9	509	10	AA636864		vr20q04.r
40	458	65.4	540	10	AA466788		vd90b09.r
41	454	64.9	473	11	W84975		mf42g01.r1
42	453.4	64.8	662	11	BF681719		602116954
43	452.2	64.6	457	10	BE654418		UI-M-AH1-
44	451	64.4	459	10	BE198421		uq78d07.y
45	450.6	64.4	533	10	AA214902		mu76d05.r

ALIGNMENTS

RESULT 1	AK004531	809 bp	mRNA	05-JUL-2001
LOCUS	Mus musculus 18 days embryo cDNA, RIKEN full-length enriched library, clone:1190011J23, full insert sequence.			
DEFINITION	Mus musculus 18 days embryo cDNA, RIKEN full-length enriched library, clone:1190011J23, full insert sequence.			
ACCESSION	AK004531			
VERSION	AK004531.1 GI:12835760			
KEYWORDS	CAP trapper.			
SOURCE	Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA, clone:1190011J23.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Methods in enzymology. 303, 19-44 (1999)			
MEDLINE	99279253			
PUBMED	10349636			
REFERENCE 2	(bases 1 to 809)			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome research. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374			
PUBMED	11042159			
REFERENCE 3	(bases 1 to 809)			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,			

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome research. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861

4 (bases 1 to 809)  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5 (bases 1 to 809)  
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,  
 Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,  
 Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F.,  
 Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,  
 Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T.,  
 Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y.,  
 Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H.,  
 Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
 Tanaka, T., Tejima, Y., Toya, T., Yamamura, F., Yasunishi, A.,  
 Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for  
 further details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues. First strand cDNA was primed with a primer  
 [5', GAGAGAGAGCGCGCGCACTCGAGCTTTTCTTTTCTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse transcriptase  
 and subsequently enriched for full-length by cap-trapper. cDNA went  
 through one round of normalization to Rot = 5.0. Second strand cDNA  
 was prepared with the primer adapter of sequence [5'  
 GAGAGAGAGAGATCAAGCTCAATTAAATTAATTAACCCGCCCC 3']. cDNA was  
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
 XhoI. Host: SOLR.

Location/Qualifiers  
 1..809  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /db\_xref="MGD:MGI:1338016"  
 /db\_xref="MGD:MGI:1892692"  
 /clone="1190011J23"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="18 days embryo"  
 181..624  
 /note="putative"  
 /codon\_start=1  
 /protein\_id="BAB23350.1"  
 /db\_xref="GI:12835761"  
 /translation="MANVQNEEWEOPLONGEDRPVGGEGHOPANNNNNNHNN  
 HNHRRCGARLPANFRWAIPIRQNDLGGDDGMFMEIMEIRKRLRELQLRNC  
 LRIUMGEUNSHDHDHDFCLPWLKSPFPEIHTVPAVALFLAFS"  
 790..795  
 /note="putative"  
 809  
 /note="putative"  
 208 a 211 c 227 g 163 t

polyA\_signal  
 polyA\_site

BASE COUNT 208 a 211 c 227 g 163 t  
 ORIGIN

Query Match 98.2% Score 687.4; DB 12; Length 809;

Best Local Similarity 99.1%; Pred. No. 2.3e-152;  
 Matches 691; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 acgagcgtctgcccagcagctcgagagctcctctcgcgcgcgcggtgacgagcgccg 60  
 DB 5 ACGAGCGTCTGGCCAGCAGCTCGAGCTCCTCTGCGCGCGGGCTGCAGCGGCCG 64  
 QY 61 agcgagcggggacagatgactggaagccgagagagagagagagagagagagag 120  
 DB 65 AGCGAGCGGGGACACATTTGACTTGGAAAGCCGACAGTCCAGCGCGGGAATTGACAG 124  
 QY 121 aggactagccgcaaggatagcccagaaatagcaaccaggagaaacaaatctcatc 180  
 DB 125 AGGACTACGCCCAAGGGATAGCCAGAAATAGCAACCAAGAAACAAAATCTCATCGG 184  
 QY 181 ccaatgctccaccagagaaagaagagctggagcagccctcgagagatgacaggaacac 240  
 DB 185 CCAATGTCCACAGGAAACGAAGAGATGAGCAGCGCCCTGGCAGATGGACAGGAAGAC 244  
 QY 241 gccctgtggagagagtgaggccaccagcctgctgcaaacacacacacacacac 300  
 DB 245 GCCCTGTGGAGAGGTGAGGGCCACACCCCTGCTGCAACAAACAACAACAACCA 304  
 QY 301 accataaacacacaccaccaggaagggccagggctcgccgacctgcccctaactcc 360  
 DB 305 ACCATAAACACACACCACCCAGAGAGGCGCAGGCTCGCCGACTTGCCCTAACTTC 364  
 QY 361 gggccattcccaacagggagatgaatgacgggttggtggagatggagatgatagaa 420  
 DB 365 GGGCCATTTCCCAACAGGCAGATGAATGACGGGTGGGTGGAGATGATATGAA 424  
 QY 421 tttccatgagagatgagagatccgagagagagagagagagagagagagagag 480  
 DB 425 TTTTCATGAGAGATGAGAGAGATCCGGAGAAAGCTTAGGAGCTTACAGCTGAGA 484  
 QY 481 gtctcgcctccttgggggagctgtctaacacacacacacacacacacacacac 540  
 DB 485 GTCTAGCATCCTTATGGGGAGCTGTCTAACCAACACACCATCACCATGATGA 544  
 QY 541 ttatccttgactcgggtcattccccctcgagatccatctgactcccgctgtagcc 600  
 DB 545 TTATGCTTGGCTCGGTCAATTCGCCCTGAGATCCATCTGTGACTCCCGCTG 604  
 QY 601 ttctcctgcatttctcgcagatgccttcaatgacccgtttgtgtagccctgtgt 660  
 DB 605 TTTTCTCGCATTTTCTTGACATGCTTTTAATGACCCGTTTGGTGAGCCCTTG 664  
 QY 661 ttccatgcatgcccagtgagggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 697  
 DB 665 TTCCATGCCATGTGCCAGGTGGGGCTTGTGTGGCCAG 701

RESULT 2  
 AK010500  
 LOCUS AK010500.1 GI:12845991  
 DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library,  
 clone:2410015K23, full insert sequence.  
 ACCESSION AK010500  
 VERSION AK010500.1  
 KEYWORDS CAP trapper.  
 SOURCE Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,  
 clone.lib:RIKEN full-length enriched mouse cDNA library  
 clone:2410015K23.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 845)  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Methods in enzymology. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE 2 (bases 1 to 845)











WashU-VHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:462548  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 507.

<b>FEATURES</b>	<b>SOURCE</b>
-----------------	---------------

```

source
1. .612
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="IMAGE:10090"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I oligo(dT) primer [5',
TGTTAGATCTCAATGGAGGCGGCCCTTTTCTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[GTGGATTCGGTACC], digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library constructed by Bob Barstead."
168 a 167 c 175 g 102 t

```

Query Match 84.3%; Score 590.4; DB 10; Length 612;  
Best Local Similarity 99.0%; Pred. No. 1.9e-129;  
Matches 594; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

**Qy** 14 cafcagctcgga~~gctcctctgcgcgcggcggttggcaggccgagcgagcggac~~ 73  
|||||  
**Db** 13 CAGCAGCTCGGAGTCTCTCTGGCGCGCGGTGGCACGGGGCCGAGGCAGCGGGAC 72

Qy 74 agattgactggaagccgagagatccaggggcagcgggaattgacaggaggactacgcgc 13

Qy 134 aagggatagcccagaatagcaaccagggaacaaaatctcatctatggccaatgtccacca 19

133 AAGGGATAGGCCCGCCAGAAATAGCAACCAAGGAAACAAAATCTCATCATGCGCAATGTCCACCA 19

194 ggaataacgaagagctggagcagccccctgcagaaatggacaggaaacacccgccctgtggagg 25

**Db** 193 GGAAACGAAGAGATGGAGCAGCCCTGCAGAATGGACAGGAAGACCGCCCTGTGGGAGG 25  
**Qy** 254 aggtgaggccaccagcctgctgcacaaacaacaaaccaaacataaaccaaa 312

Db 253 AGGTGAGGCCACCAGCCTGCTGCAACACACACACACACACCAACCAACCAATACCAACA 314

Db 313 CCACCACCGAAGGCGCAGGCTCGCGACTTGCCCTAACTTCGATGGGCCATTCCCAA 37

Qy 374 caggcagatgaatgacgggttgggtagagatgagatgatatggaatgttcattggagga 433  
|||||  
Db 373 CAGGCAGATGAATGACGGGTTGGGTGGAGATGGAGATGATGGAATGTTTCATGGAGGA 433

Qy 434 gatgagagatccggagagaagcttagggagctacagctgagaaaattgtctacgcacct 49  
 |||||  
 Db 433 GATGAGAGATCCGGAGAAGCTTAGGGAGCTACAGCTGAGAAATGTGTACGCATCCT 49  
 |||||

Qy 494 tatgggggagctgtctaaccaccacgatcaccatgatgaattctgccttatgccttgact 55  
|||||  
493 tatgggggagctgtctaaccaccacgatcaccatgatgaattctgccttatgccttgact 55  
|||||  
492 tatgggggagctgtctaaccaccacgatcaccatgatgaattctgccttatgccttgact 55  
|||||  
491 tatgggggagctgtctaaccaccacgatcaccatgatgaattctgccttatgccttgact 55  
|||||

453 TATGGGGGAGCTGTGCTAAACACACGATCACCAATGATGATTCAGCCCTTATGCCCCCTTGACT 35

Qy	554	toggtaattccccctgagatccatactgtgaactccgcgtgtagccctttcccctgcatt	613
Dd <th>553</th> <th>TCGGTCATTCCCCCCTGAGATCCATCTACTGTGACTCCGCTGAGCCCTTTCTCTCGATT</th> <th>612</th>	553	TCGGTCATTCCCCCCTGAGATCCATCTACTGTGACTCCGCTGAGCCCTTTCTCTCGATT	612

## RESULT 8

AK003294	AK003294	826 bp	mrna	HTC	05-JUL-2001
LOCUS	Mus musculus	18 days embryo	cdna, RIKEN full-length enriched library, clone:1110002F04, full insert sequence.		
DEFINITION	AK003294				
ACCESSION	AK003294.1	GI:12833872			
VERSION	CAP trapper.				
KEYWORDS	Mus musculus (strain:C57BL/6J) 18 days embryo	cdna to mrna, clone.lib:RIKEN full-length enriched mouse cdna library			
SOURCE	clone:1110002F04.				

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 826)  
Carninci, P. and Hayashizaki, Y.  
REFERENCE  
AUTHORS

**TITLE**

JOURNAL  
METHODS IN ENZYMOLOGY 303, 19-44 (1999)  
MEDLINE  
99279253  
PUBMED  
10349636

## REFERENCE

**AUTHORS** Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

**TITLE** Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

**JOURNAL** Genome research 10 (10): 1617-1630 (2000).

## MEDLINE

PUBMED  
11042159  
REFERENCE  
3 (bases 1 to 826)  
AUTHORS  
Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
Konno H., Akivama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,

**TITLE**  
RIKEN integrated sequence analysis (RISA) system—384-format  
sequencing pipeline with 384 multicapillary sequencer  
**JOURNAL**  
Genome research. 10 (11), 1757-1771 (2000)

## MEDLINE

110/68861  
4 (bases 1 to 826)  
THE RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
PUBLISHED  
AUTHORS  
REFERENCE

**TITLE**

JOURNAL  
NATURE 409, 685-690 (2001)  
REFERENCE  
5 (bases 1 to 826)  
AUTHORS  
ADACHI T, AIZAWA K,  
AKABIRA S, AKIMURA T,  
AKOYAMA H, ARAI A

REVIEWS

Hanaoki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

**TITLE** Direct Submission

**JOURNAL**

Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

COMMENT: Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

## COMMENT'T

Please visit our web site (<http://genome.qsc.riken.go.jp/>) for













```
|||||
Db 1 CGCGGGGGCTGGCAGCGGGCCGAGCGGAGCGGAGACAGATTGACTGGAAGCCGAGAGTC 60
QY 97 caggcgagcgggaattgacagaggagactagccqcaaggagtagccagagaatagcaa 156
|||||
Db 61 CAGCGCGCAGCGGGGAATTGACAGGAGGACTACGCCGCAAGGGATAGGCCCCAGAAATAGCAA 120
QY 157 ccaggaaacaaatctcatcatgcccattgtccaccaggaaacgaagagctggagcagc 216
|||||
Db 121 CCAGGAACAANAATCTCATCATGGCCCAATGTCCACCAGGAACACGAGAGATGGAGCAGC 180
QY 217 ccctgcagaatgacaggaacaccgcccctgtgggaggtgagggccaccagccctgctg 276
|||||
Db 181 CCTGTGAGAATGACAGGAAGACCGCCCTGTGGGAGGAGTGAGGGCCACCAGCCCTGCTG 240
QY 277 caaacaacaacaacaaccacaacataaaccacaaccacaacgaagagccaggtcc 336
|||||
Db 241 CAAACAACAACAACAACCACAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 300
QY 337 gccgacctgcccctaacttccgatggccattcccaacaggcagatgaatgacgggttgg 396
|||||
Db 301 GCCGACTTGCCCTTAACCTCCGATGGCCATTCCCAACAGGCAGATGAATGACGGGTGG 360
QY 397 gtggagatggagatgatggaatgttcattgagagagatgagagagatccggagaaagc 456
|||||
Db 361 GTGGAGATGGAGATGATATGGAAATGTCATGGAGGAGATGAGAGAGATCCGGAGAAAGC 420
QY 457 ttagggagctacagctgagaaattgtctacgcatccttatggggagactgtcttaaccac 516
|||||
Db 421 TTAGGGAGCTACAGCTGAGAAATGTCTACGCATCCTTATGGGGGAGCTGTCTAACCAAC 480
QY 517 acgatcaccatgatgaattctgccttatgccttgacttcggtcattccccctgagatcc 576
|||||
Db 481 ACCATCACCATGATGAATCTGCTTATGCCCTTGACTTCGGTCATTGG-CCCTGAGATCC 539
QY 577 atactgtgactcc 589
|||||
Db 540 ATACTGTGACTCC 552
```

Search completed: March 11, 2002, 16:09:50  
Job time: 10885 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2002, 16:33:03 ; Search time 1319.57 Seconds  
(without alignments)  
11139.225 Million cell updates/sec

Title: US-09-327-750D-29

Perfect score: 891

Sequence: 1 accccatccccctctat.....aataagcaatttaaaagc 891

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- GenEmbl.\*
- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.om.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.sy.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htgo\_hum.\*
- 31: em.htgo\_inv.\*
- 32: em.htgo\_rod.\*
- 33: em.htg\_hum.\*
- 34: em.htg\_inv.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	889.4	99.8	891	9	HUMOGC	M38188 Human unkno
2	889.4	99.8	891	9	AF187064	AF187064 Homo sapi
3	609.4	68.4	793	9	BC003190	BC003190 Homo sapi
4	363	40.7	854	10	AF097440	AF097440 Mus muscu
c 5	354.2	39.8	421	11	G72708	G72708 MARC 4953-4
6	312.2	35.0	700	10	AF187066	AF187066 Mus muscu
7	264.4	29.7	519	10	AF187065	AF187065 Rattus no
8	245.4	27.5	98274	9	HS71487	Z99755 Human DNA s
9	242.4	27.2	244	11	GI3550	GI3550 human STS S
10	185.2	20.8	45678	9	HSV351F8	Z70719 Human DNA s
11	158.6	17.8	69648	9	HS635G19	AL035494 Human DNA
12	141.2	15.8	1229	9	AK000959	AK000959 Homo sapi
13	141.2	15.8	1364	6	AX100231	AX100231 Sequence
14	141	15.8	43952	9	HSU10564	Z92846 Human DNA s
c 15	132.4	14.9	477	11	G35294	G35294 human STS S
c 16	122.2	13.7	35714	9	HS198P4	AL008708 Human DNA
c 17	117.2	13.2	154577	2	AL357046	AL357046 Homo sapi
c 18	117.2	13.2	154959	2	AC009282	AC009282 Homo sapi
c 19	117.2	13.2	185257	9	AL355305	AL355305 Homo DNA
20	110.8	12.4	31321	9	HSV870H8	Z70233 Human DNA s
21	110.8	12.4	41029	2	HSU08081	AL022169 Homo sapi
22	109.8	12.3	642	9	AF183416	AF183416 Homo sapi
23	109.8	12.3	828	9	AF220189	AF220189 Homo sapi
c 24	109.2	12.3	40584	9	AL133348	AL133348 Human DNA
25	108.2	12.1	791	9	AF237783	AF237783 Homo sapi
26	104	11.7	898	6	AX078272	AX078272 Sequence
c 27	102.4	11.5	744	9	AF251053	AF251053 Homo sapi
c 28	77	8.6	43952	9	HSU10564	Z92846 Human DNA s
c 29	74.6	8.4	1163	9	BC005988	BC005988 Homo sapi
c 30	71.6	8.0	504	11	G24641	G24641 human STS W
c 31	66.4	7.5	785	10	AF097439	AF097439 Mus muscu
c 32	62.4	7.0	7218	6	I66494	I66494 Sequence 14
33	59.6	6.7	40822	9	HSCU177E8	Z68694 Human DNA s
34	59.4	6.7	138872	2	AC027560	AC027560 Homo sapi
c 35	59.4	6.7	176039	2	AC073065	AC073065 Homo sapi
c 36	58.6	6.6	1208	9	AK026349	AK026349 Homo sapi
c 37	58.6	6.6	91480	9	HSJ1823F3	AL079333 Human DNA
c 38	58.4	6.6	156275	2	AL513285	AL513285 Homo sapi
39	57	6.4	196367	9	AC020740	AC020740 Homo sapi
c 40	56.6	6.4	2269	10	AF097437	AF097437 Mus muscu
c 41	56.6	6.4	186896	2	AC037486	AC037486 Homo sapi
c 42	56.2	6.3	127606	2	AL356298	AL356298 Homo sapi
c 43	56.2	6.3	194019	2	AC053464	AC053464 Homo sapi
c 44	56.2	6.3	197105	2	AL355355	AL355355 Homo sapi
c 45	56.2	6.3	199891	9	CNS00M80	AL079303 Human chr

ALIGNMENTS

RESULT 1  
HUMOGC  
LOCUS Human unknown protein from clone pHR74 mRNA, complete cds.  
DEFINITION  
ACCESSION M38188 X58942  
VERSION M38188.1 GI:189378  
KEYWORDS  
SOURCE Human ovarian granulosa cell line, cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 891)  
AUTHORS Rapp,G., Freudenstein,J., Klaudiny,J., Mucha,J., Wempe,F.,  
Zimmer,M. and Scheit,K.H.  
TITLE Characterization of three abundant mRNAs from human ovarian  
granulosa cells  
JOURNAL DNA Cell Biol. 9 (7), 479-485 (1990)  
MEDLINE 91025550  
COMMENT Draft entry and computer-readable sequence for [DNA 9, 479-485  
(1990)] kindly submitted



```
Db 1 ACCCATCCCTCCACTCTTACCGTCTCCATTTTGGTGCCTGCAAACTCTGGCAAG 60
|||||
Qy 61 aatcccggaacgaaaaatgttggtttggggaaggagagtaagggcagaaaaactga 120
|||||
Db 61 ATCCCGGGAACGAAATAATGTTGGTGTGGGGGAAGGAGGTAAAGGGAGAACTTGA 120
|||||
Qy 121 gggaggggtttaaattggggcccgtagagacgcgcggaacttctaagggtgggaaaaa 180
|||||
Db 121 GGGAGGGGCTTAATTGGAGGCCCCGTAGAGGACGCGGAACTTCTAAGGTGGGAAAAA 180
|||||
Qy 181 agaaattaaaaactcttgatatcagggtctgaatcctctggtgagacacccaagc 240
|||||
Db 181 ACGAAATTAATAATCTTTGATATCAGGCTCTGAATCCTCTGCTGAGACCAAGC 240
|||||
Qy 241 attcagctctcctcctctgttcttactgtgttcaaaagaaaaaacacagaaaaa 300
|||||
Db 241 ATTCAGTCTCTCTCTGTGCTTTGCTTACTTGTGTCAAGAAACCAACAGAAAAA 300
|||||
Qy 301 aaatctcatcgtggaataattccacaggaagaaagagatggagcagcctatgcaga 360
|||||
Db 301 AAAATCTCATGCGCAATATTCACGAGAAACGAGAGATGGAGCAGCCTATGCAGA 360
|||||
Qy 361 atggagagaagacgcgccccttgggaggggtgaaggccaccagcctgcaggaaatcgac 420
|||||
Db 361 ATGGAGAGGAAGACCCCTTTGGGAGGAGGTGAAGCCACCAGCCTGCAGGAAATCGAC 420
|||||
Qy 421 ggggacagctcgacgactgcccctaatcttcgatggccatacccaatagcagatca 480
|||||
Db 421 GGGGACAGCTCGCCGACTGCCCCCTAAATTTTCGATGGGCCATACCAATAGGAGATCA 480
|||||
Qy 481 atgatgggatgggtgagatgagatgatagaaatttcattggaggagatgagagaaa 540
|||||
Db 481 ATGATGGGATGGGTGGAGATGGAGATGATGGAATATTCATGGAGGAGATGAGAGAAA 540
|||||
Qy 541 tcagaagaaacttagggcgctgcagtgagggaattgctgcgtatccttaattgggggagc 600
|||||
Db 541 TCAGAAGAAACTTAGGGAGCTGCAGTTCAGGAATGCTGCGTATCTCTTATTCGAGGAGC 600
|||||
Qy 601 tcttaatacaccatcacatcatgatgatgatgaaatttccttatgcttgactcctgccattta 660
|||||
Db 601 TCTTAATCACCATGACCATCATGATGATGAATTTGGCTTAATGCTTGACTCTCCCATTTA 660
|||||
Qy 661 tcatgagattaaactgtgattccgcgtgtttcttttcttcttgctgatttccctaataatgc 720
|||||
Db 661 TCATGAGATTAATACTGTGATTCGCCCTGTTCTTTTTCCTTGCAATTTCTCTAATATGC 720
|||||
Qy 721 ctttactgacggtttgtgtgaaacctatgtatttccatgtgtcaagtggggtcttctg 780
|||||
Db 721 CTTTACTGATCCGTTGCTGTGAACCCATGTTATTTCCATGTGTCAAGTGGGTCTTGTG 780
|||||
Qy 781 ttgcagcttctatttgaagattgcctttgcaactcagtgatgaagtttctgtaacagtaag 840
|||||
Db 781 TTGCCAGCTCTTATTTCAGATATGCCCTTTGCACTCAGTGAAGTTTCTCTCAGCAGTAGT 840
|||||
Qy 841 ttaccocatttgcattggaataatttaaaagcccaataaagcaatttaaaaaagc 891
|||||
Db 841 TTCACCCATTTCGATGGAATAATTTAAAGCTTAATAAGCAATTTAAAAAGC 891
|||||
```

RESULT 3  
BC003190  
LOCUS  
DEFINITION  
Homo sapiens, p75NTR-associated cell death executor; ovarian granulosa cell protein (13kD), clone MGC:802 IMAGE:3357965, mRNA, complete cds.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

1 (bases 1 to 793)  
Strausberg,R.  
Direct Submission  
Submitted (13-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
BC Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcgsc.bc.ca  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAL Plate: 6 Row: k Column: 22  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7657043.  
FEATURES  
Location/Qualifiers  
1..793  
/organism="Homo sapiens"  
/db\_xref="LocusID:27018"  
/db\_xref="taxon:9606"  
/clone="MGC:802 IMAGE:3357965"  
/tissue\_type="Eye, retinoblastoma"  
/clone.lib="NIH\_MGC\_16"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
196..531  
/codon\_start=1  
/product="p75NTR-associated cell death executor; ovarian granulosa cell protein (13kD)"  
/protein\_id="AAH03190.1"  
/db\_xref="GI:13112031"  
/translation="MANHQNENEMEQPMQNGEDRPLGGGEGHQPAGNRRGQARRLA  
PNFRWAIPIRQINDMGDGDDEIFMEEMREIRKRLKRLQLRNLRLIMGELSNHHD  
HHDEFCLMP"  
BASE COUNT 225 a 170 c 205 g 193 t  
ORIGIN  
Query Match 68.4%; Score 609.4; DB 9; Length 793;  
Best Local Similarity 99.8%; Pred. No. 1.1e-157;  
Matches 610; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 281 agaaaaaacacagaaaaaaatctcatcgcaaatatttcacagaaaaagaaaga 340  
|||||  
Db 165 AGAAAAACACACAGAAAAAAATCTCATGTGGCAATATTCACCGAGAAACGAGA 224  
|||||  
Qy 341 gatggagcagcctatgcagaaatggagaggaagaccgcctttgggaggggtgaaggcca 400  
|||||  
Db 225 GATGGAGCAGCCTATGCAGAAATGGAGAGAACCCCTTTGGGAGGAGGTGAAGGCCA 284  
|||||  
Qy 401 ccagcctgcaggaaatcgacggggagcagctgcgcgacttgccttaatttcgatgggc 460  
|||||  
Db 285 CCAGCCTGCAGGAAATTCGACGGGACAGCCTCGCCGACTTGCCTTAATTTTCGATGGC 344  
|||||  
Qy 461 cataccaatagcagatcaatgatgggatgggtggagatggagatgatgaaatatt 520  
|||||  
Db 345 CATACCCAATAGCAGATCAATGATGGGATGGGTGGAGATGGAGATGATGGAATATT 404

```
QY 521 catggaggagatgagagaatcagaagaaacttagggagctgcagctgaggaattgtct 580
|||||
Db 405 CATGGAGGAGATGAGAGAAATCAGAAGAAACTTAGGGAGCTGCAGTTGAGGAATTTGTCT 464

QY 581 gcgtatccattatggggagctcttaatacacatgaccatcatgatgaattttgacctat 640
|||||
Db 465 GCGTATCCCTATGATGGGGAGCTCTTAATCACCATGACCATCATGATGAATTTTGCCTTAT 524

QY 641 gccttgacctcctgccatttatcatgatgagattaactactgtgattccgcgctgtttcttttc 700
|||||
Db 525 GCCTTGACCTCCGCCATTTATCATGAGATTAATACATGTGTGATTCCTGCTGTTCTTTTC 584

QY 701 ctgtcatttctcctaataatgcctttactgacccgtttgctgtggaacccctatgtatttcca 760
|||||
Db 585 CTGTGCATTTTCTTAATATGCCTTTACTGATCCGTTTGCTGTGAACCTATGTTATTTCCA 644

QY 761 tgtgtcaagtgggtctgtgttgccagcttctatttgaagattgcctttgacctcagtgct 820
|||||
Db 645 TGTGTCAAGTGGGTCTGTGTGTTGCCAGCTTCTATTTGAAGATTGCTTTGTCACCTCAGTGT 704

QY 821 aaqtttctgcagcagtagtttcacccatttgcattgagaaattttaaagccaaataagca 880
|||||
Db 705 AAGTTTCTGTGACGACGTAGTTTACCCATTTCATGAGAAATTTTAAAGCTAATAAAGCA 764

QY 881 atttaaaagc 891
|||||
Db 765 ATTTAAAGC 775

RESULT 4
AF097440 854 bp mRNA ROD 13-APR-1999
LOCUS Mus musculus brain expressed X-linked protein 3 (Bex3) mRNA,
DEFINITION complete cds.
ACCESSION AF097440
VERSION AF097440.1 GI:4580593
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 854)
Brown, A.L. and Kay, G.F.
Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
Hum. Mol. Genet. 8 (4), 611-619 (1999)
JOURNAL 99172070
MEDLINE Erratum: [[published erratum appears in Hum Mol Genet 1999
REMARK May;8(5):943]]
REFERENCE 2 (bases 1 to 854)
AUTHORS Brown, A.L. and Kay, G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
Location/Qualifiers
1..854
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/db_xref="dbEST:AA272375"
/chromosome="X"
/map="near Plp"
/tissue_type="pooled organs"
1..854
/gene="Bex3"
172..546
/gene="Bex3"
/codon_start=1
/product="brain expressed X-linked protein 3"
/protein_id="AAD24431.1"
/db_xref="GI:4580594"
/translation="MANVHQNDEEMEQPLONGQEDRPVGGEGHQPAAANNNNHHNHN
HNHRRGQARLAPNFRWAFPNQMNDGLGGDDMMFMFMEMREIRRKRLQLRNC
```

```
BASE COUNT 237 a 212 c 228 g 177 t
ORIGIN
LRILMAGELSNHHDDHDEFCLMP"
Query Match 40.7%; Score 363; DB 10; Length 854;
Best Local Similarity 78.8%; Pred. No. 1.6e-89;
Matches 516; Conservative 0; Mismatches 90; Indels 49; Gaps 5;
QY 284 aaaaacacacagaaaaaaatctcatcatggcacaataatctcaccagggaacgaagagat 343
|||||
Db 144 AATAGCAACACAGAAACAAATAATCTCATATGCGCAATGTCCACAGGAAACGAAGAT 203

QY 344 ggaagcagctatgcagaatgagagagagacccgcttttggagagagtgagagccacca 403
|||||
Db 204 GGAGACGCCCTTGCAGAAATGGACAGAAAGCCCTGTGGGAGGAGGTGAGGGGCCACCA 263

QY 404 gctctcagga-----atcgacggg 424
|||||
Db 264 GCCTGCTGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAGAG 323

QY 425 acaggctcgcagacttggccctaattttcgatgggccaatacccaataggcagataatga 484
|||||
Db 324 CCAGGCTCGCGACATTGCCCTTAACCTTCGGATGGGCCAATTCACAACAGCAGATGAATGA 383

QY 485 tgggatgggtgagatgagatgatatgaaatatctcatggagagagatgagagaaatcag 544
|||||
Db 384 CGGGTTGGGTGGAGATGGAGATGATGAAATGTTTCATGGAGGAGATGAGAGATCGG 443

QY 545 aagaaaacttagggagctgcagttgaggaattgtctgcgtatccttatggggagctctc 604
|||||
Db 444 GAGAAAGCTTAGGGAGCTACAGCTCAGAAATTTGCTACGCATCCTTATGGGGAGCTGTC 503

QY 605 taatcacatgaccatcatgatgaattttgccttgccttgccttgccttgccttgccttgc 663
|||||
Db 504 TAACCACCAACGATCACCATGATGAATTTGCTTATGCTTTCCTCGCATTTTCTCGCAT 563

QY 664 tgagattaactgtgattcccgctgtttctttcttcttcttcttcttcttcttcttcttct 723
|||||
Db 564 TGAGATCATCATATGATGACTCCGCTGTGACCCCTTTTCTCGCATTTTCTCGCATGCTCT 623

QY 724 tactgatcgtttgtgtgaaccctatgtattt-----ccatgtgtcaagtggtcttg 778
|||||
Db 624 TAATGACCCGTTTGGTGGTGAGCCTTGTGTATTTCATGCCATGTGCCAGGTGGGGCTTG 683

QY 779 tgttcagcgtctcttatttgaagattgccttgccttgccttgccttgccttgccttgcctt 837
|||||
Db 684 TGTGCCAGCTTCTAATTTGGAGATTGCGCTTCCCACTAGTGTGAGTTTCTGTCAACAGTG 743

QY 838 -agtttcacccatttgcagtgaaa--aatttaagccaataaagcaatttaaaaa 889
|||||
Db 744 CAGTGTGCCCATTAGCATGGAAGCAATGTCAAGTCAATAAAGCAACTGGGAATA 798

RESULT 5
G72708/c 5
LOCUS G72708 421 bp DNA STS 08-AUG-2001
DEFINITION MARC 4953-4954:991939031:1 SCF - porcine spleen Sus scrofa STS
genomic, sequence tagged site.
ACCESSION G72708
VERSION G72708.1 GI:15146738
KEYWORDS STS.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 421)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L. and
Keele, J.W.
TITLE Single nucleotide polymorphism (SNP) discovery in expressed porcine
genes
JOURNAL Unpublished (2001)
COMMENT
```

Contact: Preking BA  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4278  
Fax: 402 762 4173  
Email: freking@email.marc.usda.gov  
Primer A: GCAAAATGGTGAAACTCTACT  
Primer B: CCAGCAGCAAAATAGACG  
STS size: 500  
PCR profile:  
Hotstart: 95 degrees for 15 minutes  
Denature: 95 degrees for 30 seconds  
Anneal: 56 degrees  
Extension: 68 degrees for 2 minutes  
Cycles: 32 to 45  
Protocol:  
Template: 50-200 ng genomic DNA  
Primer: each 20 pmol  
dNTPs: each 88 uM  
Taq-Polymerase: 0.25 units (Qiagen HotStar)

Buffer: Commercially supplied Qiagen HotStar buffer

The STS is derived from PCR amplicons generated from genomic DNA, sequenced from each end using the amplification primers. The sequence does not necessarily represent the entire amplicon. Sequence derived from PolyPhred was trimmed from each end of each unique contig until five consecutive bases exceeded a quality score threshold of 20, and the next 10 bases averaged a quality score of 20 or greater. Amplicon size was estimated by agarose gel electrophoresis.

FEATURES

Location/Qualifiers  
1..421  
/organism="Sus scrofa"  
/strain="white composite, duroc, meishan, minzhu, fengjing, crossbreds"  
/db\_xref="taxon:9823"  
/sex="male and female"  
/clone\_lib="SCF - porcine spleen"  
/dev\_stage="adult"  
/note="Organ: spleen"  
<1..>421  
BASE COUNT 130 a 101 c 86 g 102 t 2 others  
ORIGIN

Query Match 39.8%; Score 354.2; DB 11; Length 421;  
Best Local Similarity 92.1%; Pred. No. 3.8e-87;  
Matches 384; Conservative 1; Mismatches 29; Indels 3; Gaps 1;  
QY 414 aatcgaggggacaggtcgcgacttgccttaatttcgatggccatacccaatagg 473  
DB 421 AATAGACGGGGACAAAGCTCGGACATGCGCCCTAATTCGATGGCCATACCAATAG 362  
QY 474 cagatcaatgatggatgggtgagatgagatgatgaaatatttcgatgagagatg 533  
DB 361 CAGATCAATGATGGATGGGTGGATGGAGATGATGGAATGTTTCATGGAGGATG 302  
QY 534 agagaatcagagaataattggggagctgcagttgaggaattgtctgcgtatccctatg 593  
DB 301 AGAGAATCAGGAGAAACTTAGGAGCTGCAGTTGAGGAATGTCTGCGTATCCTATG 242  
QY 594 gggagctctctaataccacatgacacatgatgaattttgccttgactgacctcg 653  
DB 241 GGGAGCTCTCTAATCACCATGACCATCATGATGAATTTTGGCTTATGCTTGACTCTG 182  
QY 654 ccatttatcatgagataactgtgattcccgctgtt---ttcttttctctgcatitt 710  
DB 181 CCATTTTCATGAGGTTAATCTGTGATGCCACTGTTGTYCTCTTTTTCCTTACATTTT 122  
QY 711 cctaataatgcctttactgatccggtttgtgtgaacctatggttatttccatgtgcaagt 770  
||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 121 CCTGATATGCCTTTACTGATCCATTGCTGTGAACCTTATGTAAATTCCTGTCAGT 62  
QY 771 gggcttctgttgcagcttcttattgaagattgccttgcactcagtaagtctt 827  
||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 61 GGGTCTGTGTACCAGCTTCTAATTTGGAGATTGCCTTGGCACCCCAAGTTCTTCTGTC 5  
RESULT 6  
LOCUS AF187066 700 bp mRNA ROD 11-JUN-2000  
DEFINITION Mus musculus p75NTR-associated cell death executor (Nade) mRNA,  
complete cds.  
ACCESSION AF187066  
VERSION AF187066.1 GI:8452897  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 700)  
AUTHORS Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,  
Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.  
TITLE Nade, a p75NTR-associated cell death executor, is involved in  
signal transduction mediated by the common neurotrophin receptor  
p75NTR  
JOURNAL J. Biol. Chem. 275 (23), 17566-17570 (2000)  
MEDLINE 20298829  
REFERENCE 2 (bases 1 to 700)  
AUTHORS Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,  
Hanaoka,T., Li,Y., Irie,S. and Sato,T.  
TITLE Direct Submission  
JOURNAL Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia  
University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA  
FEATURES  
Location/Qualifiers  
1..700  
/organism="Mus musculus"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
1..700  
/gene="Nade"  
/gene="Nade"  
177..551  
/note="Involved in the common neurotrophin receptor  
p75NTR-mediated signal transduction; Nade"  
/codon\_start=1  
/product="p75NTR-associated cell death executor"  
/protein\_id="AAF75131.1"  
/db\_xref="GI:8452898"  
/translation="MANVHQENEMEQLQNGQEDRPVGGEGHQHPAANNHNNHNNH  
HNHRRGGARLAPNFRWAI PNROMNDGLGGDDMEMEMREIRKRLRELQLRNC  
LRILMGLSNHHDHDEFCLMP"  
BASE COUNT 178 a 187 c 203 g 132 t  
ORIGIN  
Query Match 35.0%; Score 312.2; DB 10; Length 700;  
Best Local Similarity 78.5%; Pred. No. 1.7e-75;  
Matches 431; Conservative 0; Mismatches 73; Indels 45; Gaps 3;  
QY 284 aaaaacacagaaaaaaatctcatcgtgcgaattattccaccagaaacgaagat 343  
DB 149 AATAGCAACGAGAAACAAATCTCATGTGCCAATGTCCACCAGAAACGACAGAT 208  
QY 344 gagagcctctatgcagaatggagaggaagaccgccttggaggagggtgaaggccacca 403  
DB 209 GGAGCAGCCCTTCAGAAATGGACGAGAACCCCTCTGTGGGAGGAGTGGAGGCCACCA 268  
QY 404 gctgcagaa-----atcgacgggg 424  
DB 269 GCTGTGCAACAACAACAACAACAACCAACCAACCAACCAACCAACCAACCAAGAGG 328  
QY 425 acaggctcgcgacttgccttaatttctgatggcgcataccacaataggcagatcaatga 484  
||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||





Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> CTA-714B7 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.

VECTOR: pBeloBAC11

This sequence is the entire insert of clone CTA-714B7.

#### FEATURES

```

source
    Location/Qualifiers
    1..98274
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="22"
    /map="q12.2-13.2"
    /clone="CTA-714B7"
    /clone_lib="CIT978Sk-A2"
    109..136
    /note="14 copies 2 mer ac 100 conserved"
repeat_region
    183..258
    /note="LMCS repeat: matches 7840. 7913 of consensus"
    1177..1446
    /note="match: GSS: Em:AQ373724"
    1301..1356
    /note="L2 repeat: matches 2693. 2750 of consensus"
    2568..5038
    /note="LM1D repeat: matches 3673. 6170 of consensus"
    5036..5486
    /note="LMCE repeat: matches 2052. 2207 of consensus"
    5496..6885
    /note="LM1D repeat: matches -1. 1538 of consensus"
    6896..7185
    /note="L2 repeat: matches 2356. 2710 of consensus"
    complement(7187..7611)
    /note="match: GSS: Em:AQ566001"
    7369..7594
    /note="MIR repeat: matches 34. 256 of consensus"
    complement(7713..8092)
    7904..7978
    /note="match: GSS: Em:AQ071449"
    8086..8300
    /note="MIR repeat: matches 20. 95 of consensus"
    8154..8565
    /note="MIR repeat: matches 8. 255 of consensus"
    8549..8864
    /note="match: GSS: Em:AQ059599"
    complement(8897..9290)
    /note="match: GSS: Em:AQ343214"
    9235..9361
    /note="L2 repeat: matches 2619. 2749 of consensus"
    9378..9890
    /note="match: GSS: Em:AQ284832"
    10135..10428
    /note="AluX repeat: matches 1. 294 of consensus"
    10429..10563
    /note="AluSg/x repeat: matches 1. 135 of consensus"
    10811..10868
    /note="L2 repeat: matches 2648. 2705 of consensus"
    10819..11064
    /note="MIR repeat: matches 2. 255 of consensus"
    11112..11334
    /note="MIR repeat: matches 11. 255 of consensus"
    11869..12012
    /note="MIR repeat: matches 20. 167 of consensus"

```

```

repeat_region
    12925..13223
    /note="AluY repeat: matches 1. 300 of consensus"
    13290..13495
    /note="MER58A repeat: matches 9. 214 of consensus"
    13338..13833
    /note="match: GSS: Em:AQ568864"
    13688..13749
    /note="MIR repeat: matches 77. 139 of consensus"
    13840..13966
    /note="MIR repeat: matches 22. 147 of consensus"
    14162..14245
    /note="MIR repeat: matches 46. 136 of consensus"
    14925..15058
    /note="MIR repeat: matches 48. 191 of consensus"
    15994..16040
    /note="MIR repeat: matches 102. 144 of consensus"
    16116..16184
    /note="MER5B repeat: matches 3. 70 of consensus"
    16897..17005
    /note="MIR repeat: matches 9. 118 of consensus"
    17618..17700
    /note="MIR repeat: matches 108. 192 of consensus"
    17748..17912
    /note="MIR repeat: matches 73. 245 of consensus"
    18657..18871
    /note="MIR repeat: matches 35. 252 of consensus"
    18927..19230
    /note="AluX repeat: matches 1. 301 of consensus"
    19492..19559
    /note="MIR repeat: matches 2. 67 of consensus"
    19560..19821
    /note="AluY repeat: matches 34. 295 of consensus"
    19822..19959
    /note="MIR repeat: matches 67. 212 of consensus"
    19987..20286
    /note="AluY repeat: matches 1. 300 of consensus"
    20316..20512
    /note="MIR repeat: matches 82. 262 of consensus"
    20513..20741
    /note="LM1D repeat: matches 5970. 6224 of consensus"
    20769..21074
    /note="AluSg1 repeat: matches 1. 303 of consensus"
    21085..21207
    /note="LM2 repeat: matches 1776. 1898 of consensus"
    21212..21391
    /note="LM1D repeat: matches -6. 178 of consensus"
    21392..21471
    /note="MIR repeat: matches 3. 86 of consensus"
    22146..22441
    /note="AluX repeat: matches 1. 296 of consensus"
    22526..22610
    /note="LTR16C repeat: matches 267. 349 of consensus"
    22940..23204
    /note="HERV16 repeat: matches 378. 650 of consensus"
    23205..23514
    /note="AluSg repeat: matches 1. 311 of consensus"
    23919..24194
    /note="AluX repeat: matches 36. 311 of consensus"
    24496..24590
    /note="LTR16C repeat: matches 253. 349 of consensus"
    24871..24916
    /note="MIR repeat: matches 102. 143 of consensus"
    25378..25555
    /note="L2 repeat: matches 2330. 2519 of consensus"
    25673..25910
    /note="MIR repeat: matches 4. 255 of consensus"
    26722..26798
    /note="L2 repeat: matches 2572. 2649 of consensus"
    27040..27107
    /note="L2 repeat: matches 2489. 2500 of consensus"
    27108..27374
    /note="AluY repeat: matches 41. 304 of consensus"
    27375..27471

```



```

RESULT 10
HSV351F8      45678 bp      DNA      PRI      23-NOV-1999
LOCUS        Human DNA sequence from cosmid V351F8, between markers DXS366 and
DEFINITION   DXS87 on chromosome X contains ESTs.
ACCESSION    Z70719
VERSION      Z70719.1 GI:1261915
KEYWORDS     X.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 45678)
AUTHORS      Whiteley, M.
TITLE        Direct Submission
JOURNAL      Submitted (09-Apr-1995) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT      IMPORTANT: This sequence is the entire insert of clone V351F8. The
              true left end of clone V351F8 is at 1 in this sequence. The true
              right end of clone V351F8 is at 45678.
              V351F8 is from the human chromosome X-specific cosmid library.
FEATURES     Location/Qualifiers
              1..45678
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="X"
               /map="X"
               /clone="GHC-351F8"
               /clone_lib="SCCV"
               1..70
               /partial
               /note="Alu repeat: matches 80. .1 of consensus"
               642..930
               /note="Alu repeat: matches 1. .308 of consensus"
               1406..5179
               /note="L1 element fragment"
               5243..5276
               /note="17 copies of 2 mer 82 % conserved"
               5279..5567
               /partial
               /note="Alu repeat: matches 304. .1 of consensus"
               9975..10369
               /note="MSTA element fragment"
               10399..12017
               /note="MSTAR element fragment"
               11318..11978
               /note="THE1BR element fragment"
               11715..11846
               /note="MLTIR element fragment"
               12074..12170
               /note="MSTA element fragment"
               12227..12486
               /note="MSTA element fragment"
               12437..12502
               /note="MSTC element fragment"
               12756..13059
               /partial
               /note="Alu repeat: matches 308. .1 of consensus"
               14478..14650
               /note="THE1B element fragment"
               14491..14536
               /note="MSTA element fragment"
               14715..14831
               /note="THE1B element fragment"
               15099..15533
               /note="L1 element fragment"
               15534..15626
               /note="31 copies of 3 mer 85 % conserved"
               15566..15619
               /note="3 copies of 18 mer 98 % conserved"
               16352..16436
               /note="L1 element fragment"
               16728..16763

```

```

repeat_region 16764..17054
               /partial
               /note="Alu repeat: matches 308. .1 of consensus"
               19420..19449
               /note="15 copies of 2 mer 87 % conserved"
               complement(19537..20013)
               /note="match: 3' EST N51315 clone 283089"
               20071..20114
               /note="22 copies of 2 mer 98 % conserved"
               20073..20112
               /note="10 copies of 4 mer 100 % conserved"
               20073..20126
               /note="3 copies of 18 mer 87 % conserved"
               23088..23300
               /note="L1 element fragment"
               23311..23490
               /note="MLTIA element fragment"
               23318..23489
               /note="MLTIB element fragment"
               23362..23491
               /note="MLTIC element fragment"
               23507..24407
               /note="L1 element fragment"
               24376..24462
               /note="MSTC element fragment"
               24379..24491
               /note="MSTA element fragment"
               24688..24738
               /note="MSTA element fragment"
               24810..25016
               /note="L1 element fragment"
               25254..25421
               /note="L1 element fragment"
               25439..25696
               /note="MER25 element fragment"
               26284..26343
               /note="MLTID element fragment"
               26428..26491
               /note="MLTID element fragment"
               26465..26518
               /note="MLTIE element fragment"
               27446..27667
               /note="L1 element fragment"
               28954..29247
               /partial
               /note="Alu repeat: matches 308. .1 of consensus"
               30160..30325
               /note="2 copies of 83 mer 98 % conserved"
               32588..32875
               /note="match: 5' EST H68599 clone 239077"
               complement(32825..33230)
               /note="match: 3' EST H68239 clone 289077"
               35551..35802
               /partial
               /note="Alu repeat: matches 1. .260 of consensus"
               35804..35839
               /note="18 copies of 2 mer 83 % conserved"
               38625..38705
               /note="MLTIR element fragment"
               38747..38839
               /note="MLTIR element fragment"
               38949..39032
               /note="MLTIE element fragment"
               38965..39076
               /note="MLTID element fragment"
               9135 c 9327 g 12956 t
BASE COUNT 14260 a
ORIGIN

```

Query Match 20.8%; Score 185.2; DB 9; Length 45678;  
 Best Local Similarity 64.4%; Pred. No. 5.2e-40;  
 Matches 345; Conservative 0; Mismatches 178; Indels 13; Gaps 4;



repeat\_region /note="L1MD repeat: matches 1357. .1972 of consensus"  
11443. .11571  
/note="FLAM A repeat: matches 2. .129 of consensus"  
repeat\_region 11572. .11946  
/note="L1MD repeat: matches 977. .1357 of consensus"  
repeat\_region 11947. .12245  
/note="Aluy repeat: matches 1. .301 of consensus"  
repeat\_region 12246. .13098  
/note="L1MD repeat: matches 275. .977 of consensus"  
repeat\_region 13094. .13223  
/note="L1MD repeat: matches -17. .113 of consensus"  
repeat\_region 13523. .13820  
/note="MLTID repeat: matches 177. .505 of consensus"  
repeat\_region 13823. .13932  
/note="55 copies 2 mer ct 77% conserved"  
repeat\_region 13936. .13977  
/note="MLTID repeat: matches 201. .241 of consensus"  
repeat\_region 13978. .14240  
/note="Alusg1 repeat: matches 1. .270 of consensus"  
repeat\_region 14245. .14534  
/note="Alusc repeat: matches 1. .290 of consensus"  
repeat\_region 14535. .14721  
/note="MLTID repeat: matches 2. .201 of consensus"  
repeat\_region 14724. .15604  
/note="L1MC/D repeat: matches 4646. .5565 of consensus"  
repeat\_region 15605. .15909  
/note="Alusx repeat: matches 24. .312 of consensus"  
repeat\_region 16069. .16363  
/note="Alusx repeat: matches 1. .291 of consensus"  
repeat\_region 16482. .16641  
/note="Aludb repeat: matches 136. .295 of consensus"  
repeat\_region 16671. .16953  
/note="Alusx repeat: matches 1. .284 of consensus"  
repeat\_region 17164. .17280  
/note="L1P repeat: matches 4499. .4619 of consensus"  
repeat\_region 17326. .17377  
/note="26 copies 2 mer at 79% conserved"  
repeat\_region 17704. .17863  
/note="Alusc repeat: matches 118. .295 of consensus"  
repeat\_region 17864. .18175  
/note="Alub8 repeat: matches 1. .318 of consensus"  
repeat\_region 18176. .18298  
/note="Alusc repeat: matches 1. .118 of consensus"  
repeat\_region 18848. .18873  
/note="13 copies 2 mer tg 100% conserved"  
repeat\_region 18876. .19171  
/note="Alusg repeat: matches 1. .295 of consensus"  
repeat\_region 19183. .19232  
/note="25 copies 2 mer ac 100% conserved"  
repeat\_region 19246. .19533  
/note="Aluy repeat: matches 5. .292 of consensus"  
repeat\_region 19539. .19837  
/note="MER74B repeat: matches 61. .368 of consensus"  
repeat\_region 20016. .20194  
/note="MER74A repeat: matches 295. .487 of consensus"  
repeat\_region 20394. .20692  
/note="Alusg repeat: matches 1. .302 of consensus"  
repeat\_region 20701. .20809  
/note="LTR1 repeat: matches 682. .785 of consensus"  
repeat\_region 20877. .21061  
/note="LTR1 repeat: matches 1. .195 of consensus"  
repeat\_region 21186. .21636  
/note="LTR29 repeat: matches 3. .498 of consensus"  
repeat\_region 21609. .21737  
/note="MER34 repeat: matches 409. .543 of consensus"  
repeat\_region 21771. .22076  
/note="Aluy repeat: matches 2. .308 of consensus"  
repeat\_region 22878. .22974  
/note="MIR repeat: matches 92. .191 of consensus"  
repeat\_region 23452. .23582  
/note="FLAM\_C repeat: matches 1. .131 of consensus"  
repeat\_region 23583. .23624  
/note="21 copies 2 mer ta 76% conserved"

repeat\_region 23735. .24041  
/note="Alusg1 repeat: matches 1. .308 of consensus"  
repeat\_region 24045. .24072  
/note="14 copies 2 mer tt 89% conserved"  
repeat\_region 25274. .25582  
/note="Aludb repeat: matches 1. .289 of consensus"  
misc\_feature 25583. .25747  
/note="match: GSS B63622"  
repeat\_region 25959. .26012  
/note="HERVPH21 repeat: matches 23. .79 of consensus"  
repeat\_region 26144. .26456  
/note="Aluy repeat: matches 1. .311 of consensus"  
repeat\_region 27281. .27432  
/note="Aludb repeat: matches 137. .289 of consensus"  
repeat\_region 27436. .27507  
/note="L1M4 repeat: matches 5360. .5434 of consensus"  
repeat\_region 27536. .27600  
/note="L1M4 repeat: matches 2220. .2283 of consensus"  
repeat\_region 27601. .27975  
/note="MLT1A1 repeat: matches 3. .365 of consensus"  
repeat\_region 27976. .28273  
Query Match 17.8%; Score 158.6; DB 9; Length 69648;  
Best Local Similarity 60.1%; Pred. No. 1.3e-32;  
Matches 409; Conservative 0; Mismatches 229; Indels 43; Gaps 7;  
Qy 3 cccatccccactctctatccggtctccatcttctgctctcaaaagctctggaaagaa 62  
Db 64012 CCCCTTCCACAGCCATCTCTCTCTCATTTTGTCTGGGAGGCTTGAGATGGA 64071  
Qy 63 tcccggaacgaaatggtggttgggggagggaggtgaaaggagagaaagctggagg 122  
Db 64072 TC---TACAGGAAATGTTGGCTTTCGGAAGGAGGGCTCGATTAGGGCACCC 64127  
Qy 123 ggggggtttaaattggagggcccgtagagagcgcggaactctctaagtgggaaaaa-- 180  
Db 64128 CACAGGGCATACTGGCTTCTCAGTGTGAAAAAATAAATGTTAGAGTATAAAGTCAAGT 64187  
Qy 181 -----acgaaattaaaaatccttgatctacagggctctgaatccttgctggtcagag 232  
Db 64188 CAGGGCTCTGAATTTTAAAGGTGCTAGTAGGGCTCTGTCTCTCAGTCTTATCAGTC 64247  
Qy 233 caccagcattcagtcctctctctctgcttcttctt-----actgtgttcaagaaaaa 287  
Db 64248 CACCAAGAATTACGCCCATTTTCTCTCTCTCTCTCTAGGAGTATGGAGTCCAAGAG 64307  
Qy 288 caaccagaaaaaaatctcatcattgcaaatatccaccagaaaaaacagagatggag 347  
Db 64308 GAACTAGCGGCAACAATCTCAACGGGAAAAATGCCCAACAGAAACGAAGGGGAG 64367  
Qy 348 cag---cctatcagaatgagagagaccgcccttggagagagtgaaagccaccag 404  
Db 64368 CAGGGCTTCCACGCAAGTAAGAAGAAATCCCGCCATTTGGAGGGGTGAAGCCCAAG 64427  
Qy 405 cctgcaggaaat---cgacggggacaggctcgccgacttgcctaatcttctgattggcc 461  
Db 64428 CTTGGAGGAATATCAGCGGGGGCGAGTTAGCGCAGTTGTCCCTAATTTTCGATGGCC 64487  
Qy 462 atacccaatagcagatcaatgatgggatgggtggagatggagatgataatggaatattc 521  
Db 64488 ATACCTAATAGGCATAT-----TGAGCAACAATGAAGCAGAGATGATGTAGAAGGTTT 64541  
Qy 522 atggaggagatagagaaatcagaagaaacttagagagctcagctcagttgaggaattctcg 581  
Db 64542 GTAGGCGCATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGAGATAGGCACTATG 64601  
Qy 582 cgtatccttatggggagctctcttaatacccatgaccatcatgatgaatttttccttatg 641  
Db 64602 CGCTTCCAAA-----CTCCTGAACCTGACAAACATATGACATTTTCCCTCATA 64649  
Qy 642 ccttgactcctccattatc 662  
Db 64650 CTTGAATCCTAAAGTTTTTC 64670

```

RESULT 12.
LOCUS AK000959 1229 bp mRNA PRI 22-FEB-2000
DEFINITION Homo sapiens CDNA FLJ10097 fis, clone HEMBA1002458; weakly similar
to OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74.
ACCESSION AK000959
VERSION 1
KEYWORDS oligo capping; fis (full insert sequence);
Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
SOURCE mRNA, clone_lib:HEMBA1 clone:HEMBA1002458.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A.
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 1229)
Isogai,T. and Otsuki,T.
Direct Submission
JOURNAL Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5' & 3' end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..1229
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1002458"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/tissue_type="whole embryo, mainly head"
/notice="Cloning vector: pME18SFL3"
191..553
/notice="unnamed protein product"
/codon_start=1
/protein_id="BAA91443.1"
/db_xref="GI:7021946"
/translation="MESKEELAAANNLNGENAQQNEGEQAPONEEESRHLGGEGQ
KPGGNIRGRVRRLLVPNFRWAIPNRHIEHNEARDVDFVGMMEIKRKTREQMRHY
MRFTQPEPDNHYDFCLIP"
BASE COUNT 351 a 243 c 296 g 339 t
ORIGIN
Query Match 15.8%; Score 141.2; DB 9; Length 1229;
Best Local Similarity 68.4%; Pred. NO. 3.9e-28;
Matches 264; Conservative 0; Mismatches 98; Indels 24; Gaps 4;
QY 283 aaaaaccaccagaaataatctcatgaggaatattcaccaggagaaacgaagaga 342
DB 201 AAGAGAACTAGCGCAACAATCTCAACGGGGGAAATGCCCAACAAGAAACGAGGAG 260
QY 343 ttgagcag---cctatgcagagatgaggaagaccgccttttgaggaggtgaagcc 399
DB 261 GGGACGAGCCCCACGCGAGATGAAGAAGATCCGCCATTTGGAGGGGGTGAAGGCC 320
QY 400 accagcctgcaggaat---cgacggggacaggtcgccgaccttccccctaatcttcgat 456
DB 321 AGAAGCCTGGAGGAATATCAGCGGGGGCGAGTTAGCGGACTTGTCCCTAATTTTCGAT 380

```

```

QY 457 gggccataaccataggcagatcaatgatgggtggagatggagatgatgatgaaa 516
DB 381 GGGCCATACCTATAGGCATAT-----TGAGCACAAATGAAGCGAGAGATGATGTAGAAA 434
QY 517 tattcatggagagatgagagaaatcagaagaaacttagggagctgcagttgaggaatt 576
DB 435 GGTTTGTAGGGCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATGAGGCACT 494
QY 577 gtcgcgtatccttattgggggagctctctaatcaccatgaccatcatgatgaatttggc 636
DB 495 ATATCGCTTCCAAA-----CTCCTGAACCTGACACCACTTATGACTTTTGGC 542
QY 637 ttatgccttgactcctgccatttacc 662
DB 543 TCATACCTTGAATCCTTAAAGTTTTC 568
RESULT 13
LOCUS AX100231 1364 bp DNA PAT 02-APR-2001
DEFINITION Sequence 32 from Patent WO0119860.
ACCESSION AX100231
VERSION AX100231.1 GI:13539115
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1364)
AUTHORS Tang,Y.T., Hillman,J.L., Yue,H., Reddy,R., Lal,P., Shah,P.,
Azlmzai,Y., Baughn,M.R., Lu,D.A., Bandman,O., Shih,L.L. and
Patterson,C.
TITLE Proteins associated with cell differentiation
JOURNAL Patent: WO 0119860-A 32 22-MAR-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
1..1364
/organism="Homo sapiens"
/db_xref="taxon:9606"
/notice="Incyte ID No: 1990956CB1"
BASE COUNT 411 a 269 c 322 g 362 t
ORIGIN
Query Match 15.8%; Score 141.2; DB 6; Length 1364;
Best Local Similarity 68.4%; Pred. NO. 4e-28;
Matches 264; Conservative 0; Mismatches 98; Indels 24; Gaps 4;
QY 283 aaaaaccaccagaaataatctcatgaggaatattcaccaggagaaacgaagaga 342
DB 260 AAGAGAACTAGCGCAACAATCTCAACGGGGGAAATGCCCAACAAGAAACGAGGAG 319
QY 343 ttgagcag---cctatgcagagatgaggaagaccgccttttgaggaggtgaagcc 399
DB 320 GGGACGAGCCCCACGCGAGATGAAGAAGATCCGCCATTTGGAGGGGGTGAAGGCC 379
QY 400 accagcctgcaggaat---cgacggggacaggtcgccgaccttccccctaatcttcgat 456
DB 380 AGAAGCCTGGAGGAATATCAGCGGGGGCGAGTTAGCGGACTTGTCCCTAATTTTCGAT 439
QY 457 gggccataaccataggcagatcaatgatgggtggagatggagatgatgatgaaa 516
DB 440 GGGCCATACCTAATAGGCATAT-----TGAGCACAAATGAAGCGAGAGATGATGAGAAA 493
QY 517 tattcatggagagatgagagaaatcagaagaaacttagggagctgcagttgaggaatt 576
DB 494 GGTTTGTAGGGCAGATGATGGAATCAAGAGAAAGACTAGGGAACAGCAGATGAGGCACT 553
QY 577 gtcgcgtatccttattgggggagctctctaatcaccatgaccatcatgatgaatttggc 636
DB 554 ATATCGCTTCCAAA-----CTCCTGAACCTGACACCACTTATGACTTTTGGC 601

```

```

Qy 637 ttatgccttgactcctgccattatc 662
   || ||||| |||| |||||
Db 602 TCATACCTTGAATCCTAAAGTTTC 627

RESULT 14
HSU105G4      43952 bp      DNA      PRI      23-NOV-1999
LOCUS      Human DNA sequence from cosmid U105G4, between markers DXS366 and
DEFINITION DXS87 on chromosome X contains ESTs.
ACCESSION  Z92846
VERSION    Z92846.1 GI:1877217
KEYWORDS   X.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 43952)
REFERENCE  1
AUTHORS   Burton, J.
TITLE     Direct Submission
JOURNAL   Submitted (07-MAR-1997) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk
COMMENT   requests: clonerequest@sanger.ac.uk
            IMPORTANT: This sequence is the entire insert of clone U105G4. The
            true left end of clone U105G4 is at 1 in this sequence. The true
            right end of clone U105G4 is at 43952.
            U105G4 is from the Lawrence Livermore National Laboratory
            flow-sorted X chromosome cosmid library LLOXNC01.
FEATURES   source
            1..43952
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="X"
            /map="X"
            /clone.lib="LLOXNC01"
            /clone="LLOXNC01-U105G4"
            39..420
            /note="L1MB8 repeat: matches 520..920 of consensus"
            1121..1233
            /note="L1MA1 repeat: matches 61..170 of consensus"
            1121..1203
            /note="L1MB5 repeat: matches 61..143 of consensus"
            3155..3463
            /note="AlusX repeat: matches 1..300 of consensus"
            3465..3568
            /note="26 copies of 4 mer 88 & conserved"
            4207..4505
            /note="AluY repeat: matches 301..1 of consensus"
            <4809..>56786
            prim_transcript
            /note="multiple ESTs: match: R68943 W47079 W46779
            H26305 H00808; match: H28051 W46978 H13743 N40640 H26464;
            match: D61749 H84840 H88611 H24866 W32033; match: N28622
            H96650 N40269 N92750 R24391; match: H96654 N30984 R21973
            N20969 W78040; match: R67524 R70485 R70575 N27866 W19406;
            match: T85743"
            7275..7312
            repeat_region
            /note="19 copies of 2 mer 84 & conserved"
            7315..7612
            repeat_region
            /note="AluJo repeat: matches 302..2 of consensus"
            7661..7738
            repeat_region
            /note="MIR repeat: matches 183..261 of consensus"
            7881..7933
            repeat_region
            /note="MIR repeat: matches 118..174 of consensus"
            8874..9296
            repeat_region
            /note="L1 repeat: matches 4965..5390 of consensus"
            9151..10034
            repeat_region
            /note="L1PA2 repeat: matches 1..891 of consensus"
            11622..11659
            repeat_region
            /note="19 copies of 2 mer 100 & conserved"
            17600..17897
            repeat_region
            /note="AlusX repeat: matches 299..2 of consensus"
            18288..18693
            repeat_region
            /note="MSTA repeat: matches 3..424 of consensus"

19079..19315
/note="MIR repeat: matches 262..113 of consensus"
20482..20770
/note="AluJo repeat: matches 292..9 of consensus"
22479..22522
/note="THE1B repeat: matches 44..1 of consensus"
23653..23686
/note="17 copies of 2 mer 97 & conserved"
23726..23856
/note="AluJb repeat: matches 1..131 of consensus;
incomplete repeat"
23866..23958
/note="L1MB8 repeat: matches 83..1 of consensus"
26423..26480
/note="29 copies of 2 mer 81 & conserved"
27936..28102
/note="L1MD1 repeat: matches 627..442 of consensus"
27978..28196
/note="L1MD2 repeat: matches 572..344 of consensus"
complement(28445..28754)
/note="match: 5' EST H44148 clone 172618"
29196..29370
/note="FAM repeat: matches 175..1 of consensus"
29589..29640
/note="2 copies of 26 mer 98 & conserved"
30953..33085
/note="match: multiple ESTs: match: N25344 N62822 H09953
R41228 N52068; match: N21487 H97562 H09131 T10046; match:
H09952 H09187 R14909 N31459"
31131..31135
/note="Site of TN1000 transposon insertion"
33134..33352
/note="L1MD1 repeat: matches 304..97 of consensus"
34614..34742
/note="MIR repeat: matches 101..237 of consensus"
35476..35566
/note="MLT2D repeat: matches 1..88 of consensus"
35644..36167
/note="MLT2D repeat: matches 75..553 of consensus"
36304..36602
/note="AluY repeat: matches 299..1 of consensus"
36606..36838
/note="L1PA15 repeat: matches 870..640 of consensus"
36967..37400
/note="L1MB5 repeat: matches 490..913 of consensus"
37865..37996
/note="L1MB2 repeat: matches 285..419 of consensus"
37979..38231
/note="L1MB3 repeat: matches 428..698 of consensus"
38934..39028
/note="L1MB6 repeat: matches 472..576 of consensus"
39346..39461
/note="MIR repeat: matches 235..117 of consensus"
39800..40160
/note="MLT2FB repeat: matches 414..53 of consensus"
40760..40895
/note="AluX repeat: matches 1..136 of consensus;
incomplete repeat"
40899..42656
/note="L1 repeat: matches 3592..5390 of consensus"
42506..43526
/note="L1MA5 repeat: matches 1..1047 of consensus"

BASE COUNT  13852 a 8286 c 8337 g 13477 t
ORIGIN
Query Match      15.8%; Score 141; DB 9; Length 43952;
Best Local Similarity 54.0%; Pred. No. 8.6e-28;
Matches 494; Conservative 0; Mismatches 345; Indels 76; Gaps 7;
Oy 14 ctctataccggtcctccattttgtgcctgcaagctctgggaagaatcccggaac 73
   || || || || || || || || || || || || || || || || || || || ||
Db 15173 CCCCACTCCAGATCCTCCCTTTTGTGCTGTAGCCCTAGGGAAGGATC----TACAA 15228

```







**THIS PAGE BLANK (USPTO)**

OM of: US-09-327-750D-30 to: GenEmbl:\* out\_format : pfs  
Date: Mar 11, 2002 3:33 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=Cgnt2.1/USPTO.spool/US09327750/runat\_l1032002\_l01154\_20319/app\_query.fasta\_1.1472  
-DB=GenEmbl -QFM=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCGAPOP=6.000  
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09327750\_@CGNT\_1.1472  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-327-750D-30  
Query length: 128  
Database: GenEmbl:\*  
Database sequences: 1472140  
Database length: 341344837  
Search time (sec): 4557.230000

score\_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_ro:AF097438	+	688.00	1035.50	2.1e-49	835	AF097438 Mus musculus brain exp
gb_ro:AF051347	+	688.00	1035.47	2.1e-49	838	AF051347 Mus musculus REX-3 mRN
gb_ro:AF097437	+	683.00	1020.55	1.4e-48	2269	AF097437 Mus musculus Bex1 pro
gb_ro:BC003254	+	619.00	922.39	4.2e-43	2901	BC003254 Mus musculus, Similar
gb_ro:AF097436	+	589.50	887.68	3.6e-41	785	AF097436 Mus musculus brain exp
gb_pr:AF183416	+	453.50	684.45	7.6e-30	642	AF183416 Homo sapiens ovarian c
gb_pr:AF220189	+	453.50	682.56	9.6e-30	828	AF220189 Homo sapiens uncharact
gb_pr:HS1984	+	453.50	654.61	3.5e-28	35714	AL008708 Human DNA sequence
gb_pr:AF237783	+	450.50	678.38	1.6e-29	791	AF237783 Homo sapiens brain-exp
gb_pr:AF251053	+	450.00	678.08	1.7e-29	744	AF251053 Homo sapiens X-linked
gb_pr:AX078272	+	450.00	676.69	2.0e-29	898	AX078272 Sequence 76 from Pater
gb_pr:HSV870H8	+	450.00	650.32	6.0e-28	31321	270233 Human DNA sequence fro
gb_pr:AL133348	+	450.00	648.39	7.7e-28	40584	AL133348 Human DNA sequence
gb_hgt:HSU80B1	+	450.00	648.31	7.8e-28	41029	AL022169 Homo sapiens chromos
gb_sts:G24641	+	299.00	453.66	5.4e-17	504	G24641 human STS WI-11354, sequ
gb_pr:HSV351F8	+	172.50	229.78	0.0002	45678	270719 Human DNA sequence fro
gb_pr:HUMOGC	+	171.50	257.50	4.6e-06	891	M38188 Human unknown protein fr
gb_pr:AF187064	+	171.50	257.50	4.6e-06	891	AF187064 Homo sapiens p75NTR-a
gb_pr:BC003190	+	170.50	256.86	5.0e-06	793	BC003190 Homo sapiens, p75NTR-a
gb_pr:AK000959	+	170.00	252.86	8.3e-06	1229	AK000959 Homo sapiens cdna FLJ
gb_pr:AX100231	+	170.00	252.08	9.2e-06	1364	AX100231 Sequence 32 from Pate
gb_pr:HS635G19	+	170.00	222.88	0.0004	69648	AL035494 Human DNA sequence
gb_ro:AF187066	+	161.00	243.49	2.8e-05	700	AF187066 Mus musculus p75NTR-as
gb_ro:AF097440	+	161.00	242.01	3.3e-05	854	AF097440 Mus musculus brain exp
gb_sts:G35294	+	159.00	240.32	2.8e-05	477	G35294 human STS SHGC-37409, se
gb_ro:AF187065	+	135.00	206.57	0.0031	519	AF187065 Rattus norvegicus p75N
gb_sts:G27208	+	130.50	201.35	0.0061	421	G27208 MARC 4953-4954:991939031
gb_sts:G23964	+	100.50	157.11	1.79	372	G23964 human STS WI-15922, sequ
gb_hgt:AC022779	+	96.50	104.49	1.5e+03	197886	AC022779 Mus musculus clone
gb_hgt:AC023542	+	95.50	110.06	7.45e-84	76295	AC023542 Homo sapiens chromos
gb_pl:AF271894	+	95.00	133.72	35.85	2845	AF271894 zea mays lipoxigenase
gb_pr:AB040971	+	94.50	129.72	59.91	4408	AB040971 Homo sapiens mRNA for
gb_hgt:AP001332	+	94.00	103.47	1.7e+03	136678	AP001332 Homo sapiens chromo
gb_pl:AC079936	+	94.00	103.13	1.8e+03	143163	AC079936 Genomic Sequence FC
gb_hgt:AP003440	+	94.00	102.21	2.0e+03	162095	AP003440 Homo sapiens chromo
gb_hgt:AC087505	+	94.00	102.06	2.1e+03	165434	AC087505 Homo sapiens chromo
gb_hgt:AP002893	+	94.00	100.62	2.5e+03	200841	AP002893 Homo sapiens chromo
gb_hgt:AC028293	+	94.00	100.50	2.5e+03	204129	AC028293 Homo sapiens chromo
gb_pr:A25464	+	93.50	140.49	15.06	844	A25464 B.pertussis STX pilin ge
gb_ba:BPPIMX	+	93.50	140.48	15.08	845	Y00556 Bordetella pertussis fin

gb\_ba:BBFIMX + 93.50 137.19 22.97 1315 ! X74118 B.bronchiseptica fim  
gb\_pr:AF13843 - 93.50 137.19 22.97 1315 ! A73843 Sequence 1 from Pate  
gb\_pr:AF152500 - 93.00 132.03 44.54 2382 ! AF152500 Homo sapiens proto  
gb\_pr:AK024641 - 93.00 130.73 52.61 2837 ! AK024641 Homo sapiens cDNA:  
gb\_pr:AF217750 - 93.00 129.08 65.08 3547 ! AF217750 Homo sapiens proto

seq\_name: gb\_ro:AF097438

seq\_documentation\_block:  
LOCUS AF097438 835 bp mRNA ROD 11-APR-1999  
DEFINITION Mus musculus brain expressed X-linked protein 1 (Bex1) mRNA,  
complete cds.

ACCESSION AF097438

VERSION AF097438.1 GI:4580589

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 835)

AUTHORS Brown,A.L. and Kay,G.F.

TITLE Bex1, a gene with increased expression in parthenogenetic embryos,

is a member of a novel gene family on the mouse X chromosome

JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)

MEDLINE 99172070

REMARK Erratum: [[published erratum appears in Hum Mol Genet 1999

REFERENCE May:8(5):943]]

AUTHORS 2 (bases 1 to 835)

Brown,A.L. and Kay,G.F.

TITLE Direct Submission

JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of

Medical Research, Herston Rd, Brisbane, Qld 4029, Australia

FEATURES

Location/Qualifiers

1..835

/organism="Mus musculus"

/db\_xref="taxon:10090"

/chromosome="X"

/map="near plp"

/tissue\_type="blastocyst"

1..835

/gene="Bex1"

205..591

/gene="Bex1"

/codon\_start=1

/product="brain expressed X-linked protein 1"

/protein\_id="AAD24429.1"

/db\_xref="GI:4580590"

/translation="MESKDGQVKNLNMENDHOKKEEKPDQTIIRREPAVALISEAG

KNCAPRGRRFRVRQPIAHYRWDLMQRVGEPQGRMRRENVQRFGGDVRQLMEKLRER

QLSHSLRAVSTDPHHDHDFCLMP"

BASE COUNT 238 a 184 c 246 g 167 t

ORIGIN

alignment\_scores:

Quality: 688.00 Length: 128

Ratio: 5.375 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-30 x AF097438 ..

Align seg 1/1 to: AF097438 from: 1 to: 835

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspH1 17

|||||

205 ATGCAGTCCAAAGATCAAGCGTGAAATCTCAACATGGAGATGACCA 254

|||||

17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArg 34

|||||

255 TCAGAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 304

|||||

34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50

```
|||||
305 AGCCAGCTGTGGCCCTCATCTCCAGGCTGCAAAAACCTGCGCCCTAGA 354
51 GlyGlyArgArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
355 GGAGGTGCGAGCGGTTCCGGGTTTCGGCAGCCCTCATCTCACTATAGATG 404
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGluA 84
|||||
405 GGACCTCATGCAGAGGTTGGGAGCCCCAGGAGGATGAGAGAGAGA 454
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||
455 ACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTCATGTGAGAAAGCTGAGG 504
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHI 117
|||||
505 GAAAGGCAGCTGAGCCACAGCCTGCGGGGTTAGCACTACCCGCGCTCA 554
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
555 TCATGACCACCATGATGAGTTTGCCTCATGCCCC 588
```

seq\_name: gb\_ro:AF051347

```
seq_documentation_block:
LOCUS AF051347 838 bp mRNA ROD 01-OCT-1998
DEFINITION Mus musculus REX-3 mRNA, complete cds.
ACCESSION AF051347
VERSION AF051347.1 GI:3510642
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 838)
Faria,T.N., LaRosa,G.J., Willen,E., Liao,J. and Gudas,L.J.
Characterization of genes which exhibit reduced expression during
the retinoic acid-induced differentiation of F9 teratocarcinoma
cells: involvement of cyclin D3 in RA-mediated growth arrest
Mol. Cell. Endocrinol. 143 (1-2), 155-166 (1998)
99021197
2 (bases 1 to 838)
Faria,T.N., LaRosa,G., Willen,E., Liao,L. and Gudas,L.J.
Direct Submission
Submitted (26-FEB-1998) Pharmacology, Cornell University Medical
College, 1300 York Avenue, New York, NY 10021, USA
Location/Qualifiers
1..838
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="F9 teratocarcinoma"
229..579
/note="retinoic acid reduced gene expression"
/codon_start=1
/product="REX-3"
/protein_id="AAC61929.1"
/db_xref="GI:3510643"
/translaton="MENDHKKKEEKEKPOPTIRREPAVALISEAGKNCAPRGGRRRF
RVRPQIAHYRDLQMRVGEPOQREENVORFGDVRQLMEKRLRQLSHSLRAVSTD
PPHHDDHDFCLMP"
```

CDS

BASE COUNT 238 a 181 c 251 g 168 t

ORIGIN

alignment\_scores:

Quality: 688.00 Length: 128

Ratio: 5.375 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-327-750D-30 x AF051347

```
Align seg 1/1 to: AF051347 from: 1 to: 838
1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHI 17
|||||
193 ATGAGTCCAAAGATCAAGCGGTGAAAAATCTCAACATGGAGATGACCA 242
17 sGlnLysLysGluGluLysGluGluLysProGlnAspThrIleArgArg 34
|||||
243 TCAGAAAAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 292
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
|||||
293 AGCCAGCTGTGGCCCTCATCTCCAGGCTGCAAAAACCTGCGCCCTAGA 342
51 GlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
|||||
343 GGAGGTGCGAGCGGTTCCGGGTTTCGGCAGCCCTCATCTCACTATAGATG 392
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGluA 84
|||||
393 GGACCTCATGCAGAGGTTGGGAGCCCCAGGAGGATGAGAGAGAGA 442
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||
443 ACGTACAGAGGTTTGGGGGTGATGTGAGACAGCTCATGTGAGAAAGCTGAGG 492
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHI 117
|||||
493 GAAAGGCAGCTGAGCCACAGCCTGCGGGGTTAGCACTACCCGCGCTCA 542
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
543 TCATGACCACCATGATGAGTTTGCCTCATGCCCC 576
```

seq\_name: gb\_ro:AF097437

```
seq_documentation_block:
LOCUS AF097437 2269 bp DNA ROD 05-AUG-1999
DEFINITION Mus musculus Bex1 protein (Bex1) gene, complete cds.
ACCESSION AF097437
VERSION AF097437.1 GI:5702155
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2269)
Brown,A.L. and Kay,G.F.
Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
Hum. Mol. Genet. 8 (4), 611-619 (1999)
99172070
MEDLINE 10072429
PUBMED 10072429
2 (bases 1 to 2269)
Brown,A.L. and Kay,G.F.
Direct Submission
TITLE
JOURNAL
JOURNAL
SUBMITTED (07-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4030, Australia
Location/Qualifiers
1..2269
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"
/chromosome="X"
/map="near Pip"
/tissue_type="liver"
join(409..509,866..942,1263..1873)
/gene="Bex1"
/product="Bex1 protein"
409..1873
/gene="Bex1"
/note="expressed in brain; X-linked"
1268..1654
CDS
```

```

/gene="Bex1"
/codon_start=1
/product="Bex1 protein"
/protein_id="RAD47168.1"
/db_xref="GI:5702156"
/translation="MESKDDQGVKKLNMMENDHOKKEEKQDDTIRREPAVALTSEAG
KNCAPRGRRRFRVQPTAHYRWDLQMRVGEPOGRMREENVOREGGDVRQLMEKLRIR
QLSHSLRAVSTDPPHDDHDEFCLMP"
BASE COUNT      563 a 554 c 705 g 447 t
ORIGIN

```

```

alignment_scores:
  Quality: 683.00      Length: 128
  Ratio: 5.378         Gaps: 0
  Percent Similarity: 99.219      Percent Identity: 99.219

```

## alignment\_block:

US-09-327-750D-30 x AF097437 ..

Align seg 1/1 to: AF097437 from: 1 to: 2269

```

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHi 17
1268 ATGGAGTCCAAAGATCAAGGCGTGAATAATCTCAACATGAGAGTACCA 1317
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArg 34
1318 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1367
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
1368 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACCTGTGCACCTAG 1417
51 GlyGlyArgArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67
1418 GGAGGTGCGAGGGGTTCGGGGTTCGGCAGCCCATCGCTCACATATAGTG 1467
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGluA 84
1468 GGACCTGTATCAGAGGGTTCGGGAGCCCGAGGAGGAGGAGGAGGAGG 1517
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
1518 ACGTACAGAGGTTTGGGGTGTGTGAGACAGCTCATGGAGAGCTGAGG 1567
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
1568 GAAAGGCAGCTGAGCCACAGCCCTCGCGGGCTTAGCACCTGACCCGCTCA 1617
117 sHisAspHisHisAspGluPheCysLeuMetPro 128
1618 TCATGACCACCATGATGAGTGTTCCTTCATGCCCC 1651

```

seq\_name: gb\_ro:BC003254

```

seq_documentation_block:
LOCUS BC003254      2901 bp      mRNA      ROD      20-FEB-2001
DEFINITION Mus musculus, Similar to dentatorubral pallidolusian atrophy,
clone MGC:5758, mRNA, complete cds.
ACCESSION BC003254
VERSION BC003254.1 GI:13096906
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2901)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

```

## REMARK COMMENT

USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: [Robert\\_Strausberg@nih.gov](mailto:Robert_Strausberg@nih.gov)  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [villalon@bcm.tmc.edu](mailto:villalon@bcm.tmc.edu)  
 Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,  
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 9 Row: 0 Column: 6.

## FEATURES source

```

1..2901
  Location/Qualifiers
    organism="Mus musculus"
    db_xref="taxon:10090"
    clone="MGC:5758 IMAGE:3500522"
    tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
    ductal carcinoma, 5 month old virgin mouse."
    clone_lib="NCL_CGAP_Mam6"
    lab_host="DH10B"
    note="Vector: pCMV-SPORT6"
    755..2359
    /codon_start=1
    /product="Similar to dentatorubral pallidolusian atrophy"
    /protein_id="AAH03254.1"
    /db_xref="GI:13096907"
    /translation="MVPQFLLTPLTLPGLPKPIHVHTPLHPRAKHEHTPSHQSLSQSP
    PPQLPFLSPSPAATTGPTATQIQEPAEVEYPPSPVPKSPSPKVPKVVDPVS
    HASASARENKHLDRGNSCADSLYFVPLEGSKLAKRADLVKVRREAEQARAREKE
    REREREREKEREREKERSVKLAQSGRAPVPCSLGVPVHPFPFPGSAVATVPP
    YLGEDTPALRTLSEYARHVASPGNNRHPIFFVPLGAVDPGLLGLYNVALYSSDPAARE
    REREARDLRLKPPGEVKPSELEPLHGVPGGLDPFPRHGGALQPGPPGLHPFP
    FPHSLPLERLERLAAAGPALRPMDSYARERAAERAAALGNQDLARLQMLNVT
    PHHHQSHIHSHLHQDDIAHAASVSHPLIDPLASGSHLTRIPYPAGTLPNPLPH
    PLHNEVLRHOLFAPYRDLPASLSAPMSAAHQLAQMAHQSAELQRLALEQQQWLHAH
    HPLHSVPLPAQEDYXSHLKESDKPL"
BASE COUNT      623 a 958 c 791 g 529 t
ORIGIN

```

```

alignment_scores:
  Quality: 619.00      Length: 118
  Ratio: 5.291         Gaps: 0
  Percent Similarity: 99.153      Percent Identity: 99.153

```

## alignment\_block:

US-09-327-750D-30 x BC003254

Align seg 1/1 to: BC003254 from: 1 to: 2901

```

1 MetGluSerLysAspGlnGlyValLysAsnLeuAsnMetGluAsnAspHi 17
223 ATGGAGTCCAAAGATCAAGGCGTGAATAATCTCAACATGAGAGTACCA 272
17 sGlnLysLysGluGluLysGluLysProGlnAspThrIleArgArg 34
273 TCAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 322
34 luProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaProArg 50
323 AGCCAGCTGTGGCCCTGACCTCCGAGGCTGGCAAAACCTGTGCACCTAG 372
51 GlyGlyArgArgArgPheArgValArgGlnProIleAlaHisTyrArgTr 67

```

```

373 GGAGGTCCGAGCGGTTCCGGGTTCCGAGCCCATCGCTCACTATAGATG 422
67 pAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluGluA 84
|||||
423 GGACCTGATGACAGAGGTTGGGAGCCCGAGGAGGATGAGAGGAGA 472
84 snValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeuArg 100
|||||
473 ACGTACAGAGGTTTGGGGGTGATGTGACAGAGCTCATGGAGAGCTGAGG 522
101 GluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProProHi 117
|||||
523 GAAAGGAGCTGAGCCACAGCTCGCGGCTTAGCAGCTGACCCGCGCTCA 572
117 sHis 118
|||||
573 TCAT 576

```

seq\_name: gb\_ro:AF097439

```

seq_documentation_block:
LOCUS AF097439 785 bp mRNA ROD 13-APR-1999
DEFINITION Mus musculus brain expressed X-linked protein 2 (Bex2) mRNA,
complete cds.
ACCESSION AF097439
VERSION AF097439.1 GI:4580591
KEYWORDS house mouse,
SOURCE Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Bex1, a gene with increased expression in parthenogenetic embryos,
is a member of a novel gene family on the mouse X chromosome
JOURNAL Hum. Mol. Genet. 8 (4), 611-619 (1999)
MEDLINE 99172070
REMARK Erratum:([published erratum appears in Hum Mol Genet 1999
May;8(5):943])
REFERENCE
2 (bases 1 to 785)
AUTHORS Brown,A.L. and Kay,G.F.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES
Location/Qualifiers
1..785
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="dbEST:w48832"
/chromosome="X"
/map="near Plp"
/dev_stage="embryo; 15.5 dpc"
1..785
/gene="Bex2"
Erratum:8(5):943]
139..528
/gene="Bex2"
/codon_start=1
/product="brain expressed X-linked protein 2"
/protein_id="AAD24430.1"
/db_xref="GI:4580592"
/translation="MESKVEQGVKNLMNNDHQEKKEEKPODASKRDPTVALPFEA
GDYVPRGGRFRFRVROPIVHYRWDLMHRVGEQGRMRREENVQRGDDVRLMEKLE
RQLSHSLRAVSTDPPHDDHDFCLMP"

```

```

BASE COUNT 212 a 163 c 235 g 175 t
ORIGIN
alignment_scores:
Quality: 589.50 Length: 129
Ratio: 4.912 Gaps: 1
Percent Similarity: 93.023 Percent Identity: 86.822

```

alignment\_block:

us-09-327-750D-30 x AF097439 ..

Align seg 1/1 to: AF097439 from: 1 to: 785

```

1 MetGluSerLys...AspGlnGlyValLysAsnLeuMetGluAsnAs 16
|||||
139 ATGGAGTCCAAAGTGGAAACAAGCGGTGAAAAATCTCAACATGGAGAAATGA 188
16 pHisGlnLysLysGluGluLysGluGluLysProGlnAspThrIleArgA 33
|||||
189 CCATCAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 238
33 rgGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaPro 49
|||||
239 GGGATCCGATTGTGGCCCTGCCTTTCGAAGCTGGAGACTACTAGTGCCT 288
50 ArgGlyGlyArgArgArgPheArgValArgGlnProIleAlaHisTyrAr 66
|||||
289 AGAGGAGGTCGCAGCGGTTCCGGGTTCCGCGAGCCCATCGTGCACATACAG 338
66 gTrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluG 83
|||||
339 ATGGGACCTCATCATAGGTTGGGAGGCGCCAGGAGGAGGAGGAGGAGG 388
83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99
|||||
389 AGAAGCTACAGAGGTTTGGGATGATGTGAGACAGCTCATGGAGAGCTG 438
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
|||||
439 AGGAAAGGAGCAGCTGAGCCACAGCCCTCGCGGGCGTTAGCACTGACCCGCC 488
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
489 TCATCATGACCACCATCATGAGTTTGGCTTATGTCCTTATGTCCT 525

```

seq\_name: gb\_pt:AF183416

```

seq_documentation_block:
LOCUS AF183416 642 bp mRNA PRI 02-SEP-2000
DEFINITION Homo sapiens ovarian granulosa cell 13.0 kDa protein hGR74 homolog
mRNA, complete cds.
ACCESSION AF183416
VERSION AF183416.1 GI:9963770
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Peng,Y., Qian,b., Fu,Y., Xu,S., Han,Z., Fu,G. and Chen,Z.
TITLE A novel gene expressed in human adrenal gland
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 642)
AUTHORS Peng,Y., GU,W., Huang,C., Xu,S., Han,Z., Fu,G. and Chen,Z.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1999) Chinese National Human Genome Center at
Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, P. R. China
FEATURES
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="adrenal gland"
3..380
/feature="hGR74-h protein"
/evidence=not_experimental
/product="ovarian granulosa cell 13.0 kDa protein hGR74
homolog"
/protein_id="AAG09685.1"
/db_xref="GI:9963771"

```

/translation="MESKEKRAVNSLSMENANOENKEEQVANKGEPLALPLDAGEYC  
VPRGNRRFRVQPILOYRWMHRLGEPQARMRENMERIGEEVRLMEKLRKQLS  
HSLRAVSTDPHPHHDHDFCLMP"  
BASE COUNT 204 a 118 c 157 g 163 t  
ORIGIN

alignment\_scores:  
Quality: 453.50 Length: 129  
Ratio: 3.943 Gaps: 3  
Percent Similarity: 89.147 Percent Identity: 68.992

alignment\_block:  
US-09-327-750D-30 x AF183416 ..  
Align seg 1/1 to: AF183416 from: 1 to: 642

1 MetGluSerLysAsp...GlnGlyValLysAsnLeuAsnMetGluAsnAs 16  
3 ATGGAGTCCAAAGAGAAACGACGAGTAAACAGTCTCAGCATGGAATAATGC 52  
16 pHisGlnLysLysGluGluLysProGlnAspThrIleArgA 33  
53 CAACCAAGAAATGAAGAAAGAG.....CAAGTTGCTATAATAAG 93  
33 rgGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaPro 49  
94 GGGAGGCC...TTGGCCCTCCCTTTGGATGCTGGTGAATACTGTGTCCT 140  
50 ArgGlyValArgArgPheArgValArgGlnProIleAlaHisTyrAr 66  
141 AGAGAAATCGTAGGCGGTTCGCGTTAGCAGCCCATCTGCAGTATAG 190  
66 gTrpAspLeuMetGlnArgValGlyValGluProGlnGlyArgMetArgGlu 83  
191 ATGGGATATGATCATAGTCTGGAGAACACACAGGCAAGGATGAGAGAAG 240  
83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99  
241 AGAATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGAAGAGCTG 290  
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116  
291 AGGGAAGACAGTTGATGATCATAGTCTGCGGCGAGTCAGCACTGACCCCC 340  
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128  
341 TCACCATGACCATCATGATGAGTTTGGCTTATGCCCC 377

seq\_name: gb\_pr:AF220189

seq\_documentation\_block:  
LOCUS AF220189 828 bp mRNA PRI 04-MAY-2000  
DEFINITION Homo sapiens uncharacterized hypothalamus protein HBEX2 mRNA,  
complete cds.  
ACCESSION AF220189  
VERSION AF220189.1 GI:7689028  
KEYWORDS  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.  
TITLE A novel gene expressed in human hypothalamus  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 828)  
AUTHORS Xiao,H., Song,H., Gao,G., Ren,S., Chen,Z. and Han,Z.  
TITLE Direct Submission  
JOURNAL Submitted (30-DEC-1999) Chinese National Human Genome Center at  
Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,  
Shanghai 201203, P. R. China  
FEATURES Location/Qualifiers

source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="hypothalamus"  
167..544  
/codon\_start=1  
/product="uncharacterized hypothalamus protein HBEX2"  
/protein\_id="AAF67654.1"  
/db\_xref="GI:7689029"  
/translation="MESKEKRAVNSLSMENANOENKEEQVANKGEPLALPLDAGEYC  
VPRGNRRFRVQPILOYRWMHRLGEPQARMRENMERIGEEVRLMEKLRKQLS  
HSLRAVSTDPHPHHDHDFCLMP"  
BASE COUNT 256 a 170 c 221 g 181 t  
ORIGIN

alignment\_scores:  
Quality: 453.50 Length: 129  
Ratio: 3.943 Gaps: 3  
Percent Similarity: 89.147 Percent Identity: 68.992

alignment\_block:  
US-09-327-750D-30 x AF220189 ..  
Align seg 1/1 to: AF220189 from: 1 to: 828

1 MetGluSerLysAsp...GlnGlyValLysAsnLeuAsnMetGluAsnAs 16  
167 ATGGAGTCCAAAGAGAAACGACGAGTAAACAGTCTCAGCATGGAATAATGC 216  
16 pHisGlnLysLysGluGluLysProGlnAspThrIleArgA 33  
217 CAACCAAGAAATGAAGAAAGAG.....CAAGTTGCTATAATAAG 257  
33 rgGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaPro 49  
258 GGGAGGCC...TTGGCCCTCCCTTTGGATGCTGGTGAATACTGTGTCCT 304  
50 ArgGlyValArgArgPheArgValArgGlnProIleAlaHisTyrAr 66  
305 AGAGAAATCGTAGGCGGTTCGCGTTAGCAGCCCATCTGCAGTATAG 354  
66 gTrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGlu 83  
355 ATGGGATATGATCATAGTCTGGAGAACACACAGGCAAGGATGAGAGAAG 404  
83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99  
405 AGAATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGAAGAGCTG 454  
100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116  
455 AGGGAAGACAGTTGATGATCATAGTCTGCGGCGAGTCAGCACTGACCCCC 504  
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128  
505 TCACCATGACCATCATGATGAGTTTGGCTTATGCCCC 541

seq\_name: gb\_pr:HS198P4

seq\_documentation\_block:  
LOCUS HS198P4 35714 bp DNA PRI 16-DEC-2000  
DEFINITION Human DNA sequence from clone RPI-198P4 on chromosome Xq22 Contains  
a CpG Island, complete sequence.

ACCESSION AL008708  
VERSION AL008708.4 GI:4826445  
KEYWORDS HTG: CpG Island.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 35714)  
AUTHORS Corby,N.

TITLE  
JOURNAL

Direct Submission  
Submitted (21-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerequest@sanger.ac.uk

## COMMENT

On May 13, 1999 this sequence version replaced gi:4582117.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.

This sequence was generated from part of bacterial clone contigs of  
human chromosome X, constructed by the Sanger Centre Chromosome X  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/ChrX>

During sequence assembly data is compared from overlapping clones.  
where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at

[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep/RPI-198P4](http://www.sanger.ac.uk/projects/C_elegans/wormpep/RPI-198P4) is  
from the library RPCI-1 constructed at the Roswell Park Cancer  
Institute by the group of Pieter de Jong. For further details see  
<http://bacpac.med.buffalo.edu/>

VECTOR: PCYPAC2  
IMPORTANT: This sequence is not the entire insert of clone  
RPI-198P4. It may be shorter because we only sequence overlapping  
sections once, or longer because we arrange for a small overlap  
between neighbouring submissions.

The true left end of clone LLOXNC01-221P2 is at 35615 in this  
sequence. The true right end of clone LLOXNC01-73E8 is at 100 in  
this sequence.

## FEATURES

## source

Location/Qualifiers  
1..35714

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"

/map="q22"  
/clone="RPI-198P4"

/clone\_lib="RPCI-1"  
1..258

repeat\_region  
1..258  
/note="L1P3 repeat: matches 5666. 5925 of consensus"  
256..537  
repeat\_region  
/note="L1P3 repeat: matches 5391. 5672 of consensus"  
1491..1749  
repeat\_region  
/note="AluSc repeat: matches 39. 297 of consensus"  
2182..3810  
repeat\_region  
/note="L1P3 repeat: matches 4533. 6150 of consensus"  
5169..6687  
repeat\_region  
/note="L1MA2 repeat: matches 4798. 6308 of consensus"  
6976..7665  
repeat\_region  
/note="L1MA2 repeat: matches 4113. 4798 of consensus"  
8039..9213  
repeat\_region  
/note="L1MA2 repeat: matches 2945. 4113 of consensus"  
9523..10437  
repeat\_region  
/note="L1MA2 repeat: matches 2016. 2945 of consensus"  
10736..11494  
repeat\_region  
/note="L1MA2 repeat: matches 1262. 2016 of consensus"  
11807..12251  
repeat\_region  
/note="L1MA2 repeat: matches 310. 755 of consensus"  
12259..12331  
repeat\_region  
/note="HERVFB21 repeat: matches 5. 77 of consensus"  
12332..12796  
repeat\_region  
/note="L1R21A repeat: matches 1. 490 of consensus"  
16162..16371  
repeat\_region  
/note="105 copies 2 mer tt 55% conserved"  
17121..17424  
repeat\_region  
/note="MER33 repeat: matches 1. 322 of consensus"

repeat\_region  
18259..18290  
/note="16 copies 2 mer ca 87% conserved"  
18505..18672  
repeat\_region  
/note="MER31A repeat: matches 1. 181 of consensus"  
19478..19814  
repeat\_region  
/note="MER2 repeat: matches 1. 345 of consensus"  
21061..21101  
repeat\_region  
/note="HERVFB21 repeat: matches 42. 85 of consensus"  
21247..21781  
repeat\_region  
/note="MLT1 repeat: matches 44. 568 of consensus"  
24465..24492  
misc\_feature  
/note="Single clone region  
Single clone region"  
24531..24562  
misc\_feature  
/note="Weak data  
Weak data"  
24563  
misc\_feature  
/note="Single clone region  
Single clone region"  
24711..24865  
misc\_feature  
/note="region covered by subclones from a PAC DNA PCR  
product only at 12x coverage."  
region covered by subclones from a PAC DNA PCR product only  
at 12x coverage."  
24758..25460  
misc\_feature  
/note="CpG island"  
/evidence="not\_experimental"  
24875..24940  
repeat\_region  
/note="33 copies 2 mer cc 66% conserved"  
26795..26926  
repeat\_region  
/note="66 copies 2 mer ct 63% conserved"  
27088..27133  
repeat\_region  
/note="23 copies 2 mer tg 100% conserved"  
27461..27777  
repeat\_region  
/note="L1MB8 repeat: matches 5851. 6171 of consensus"  
27802..28232  
repeat\_region  
/note="L1MC/D repeat: matches 5192. 5403 of consensus"  
28536..28930  
repeat\_region  
/note="L1MC/D repeat: matches 4802. 5192 of consensus"  
29240..29341  
repeat\_region  
/note="TIGER1 repeat: matches 484. 587 of consensus"  
29603..30740  
repeat\_region  
/note="TIGER1 repeat: matches 587. 1757 of consensus"  
31027..31365  
repeat\_region  
/note="TIGER1 repeat: matches 1757. 2105 of consensus"  
31664..31691  
repeat\_region  
/note="TIGER1 repeat: matches 2105. 2132 of consensus"  
31809..32015  
repeat\_region  
/note="TIGER1 repeat: matches 2132. 2246 of consensus"  
32323..32389  
repeat\_region  
/note="TIGER1 repeat: matches 2246. 2313 of consensus"  
34353..34636  
repeat\_region  
/note="MLT1-INTERNAL repeat: matches 607. 889 of  
consensus"  
35149..35376  
repeat\_region  
/note="MLT1-INTERNAL repeat: matches 1001. 1233 of  
consensus"  
35523..35670  
repeat\_region  
/note="MLT1 repeat: matches 8. 159 of consensus"

BASE COUNT 9708 a 7094 c 6805 g 12107 t  
ORIGIN

## alignment\_scores:

Quality: 453.50 Length: 129  
Ratio: 3.943 Gaps: 3  
Percent Similarity: 89.147 Percent Identity: 68.992

## alignment\_block:

US-09-327-750D-30 x HS198P4/rev ..

Align seg 1/1 to reverse of: HS198P4 from: 1 to: 35714



```

1 MetGluSerLysAsp...GlnGlyValLysAsnLeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24358 ATGGAGTCCAAAGAGAAACAGCAGTAAACAGTCTCAGCATGGAATATGC 24309

16 pHISGlnLysLysGluGluLysGluLysProGlnAspThrIleArgA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24308 CAACCAAGAAATGAAGAAAGAG.....CAAGTTGCTTAATAAAG 24268

33 rgGluProAlaValAlaLeuLysSerGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24267 GGGAGGCC...TTGGCCCTCCCTTTGGATGCTGGTGAATACTGTGTGCCT 24221

50 ArgGlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24220 AGAGGAATCGTAGCGGTTCCGCGTTAGCAGCCCATCTGCAGTATAG 24171

66 gTrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluG 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24170 ATGGGATATGATGCATAGGCTTGGAGAACCCACAGGCAAGGATGAGAGAAG 24121

83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24120 AGAATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGAAAGCTG 24071

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24070 AGGGAAGCAGTTGAGTCATAGTCTCGCGGAGTCAGCAGCTGACCCCC 24021

116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24020 TCACCATGACCATCATGATGAGTTTGGCTTATGCC 23984

```

seq\_name: gb\_pr:AF237783

```

seq_documentation_block: 791 bp mRNA PRI 02-SEP-2000
LOCUS AF237783
DEFINITION Homo sapiens brain-expressed protein BEX1 (BEX1) mRNA, complete
          cds.
ACCESSION AF237783
VERSION AF237783.1 GI:9963898
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Yang,Q.S., Ying,K., Xie,Y. and Mao,Y.M.
TITLE A Novel Human X-linked Brain Expressed Protein
JOURNAL Unpublished
REFERENCE
AUTHORS Mao,Y.M., Xie,Y., Yang,Q.S., Wu,H., Lin,S. and Ying,K.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2000) Genetic Research Institute, Fudan
          University, 220 Handan Rd., Shanghai 200433, P.R.China
FEATURES
          Location/Qualifiers
            source
              1..791
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="X"
                /map="between DXS990 and DX1059 by RH (Stanford G3)"
                1..791
                /gene="BEX1"
                172..549
                /gene="BEX1"
                /codon_start=1
                /product="brain-expressed protein BEX1"
                /protein_id="AAG09752.1"
                /db_xref="GI:9963898"

```

```

gene
  1..791
  172..549
  /gene="BEX1"
  /codon_start=1
  /product="brain-expressed protein BEX1"
  /protein_id="AAG09752.1"
  /db_xref="GI:9963898"
CDS
  1..791
  172..549
  /gene="BEX1"
  /codon_start=1
  /product="X-linked protein"
  /protein_id="AAK34943.1"
  /db_xref="GI:13625168"
  /translation="MESKEKLAVNSLSEMANOENKEOVANKGEPLALPLDAGEYCV
  VPGNRRFRVROPILQYRWDMHRLGEPQNRNENRIGEGVQLMEKLRKQLS
  HSLRAVSTDPHPHDDHDFCLMP"
BASE COUNT 214 a 172 c 218 g 187 t
ORIGIN

```

alignment\_scores:

Quality: 450.50 Length: 129  
Ratio: 3.952 Gaps: 3  
Percent Similarity: 88.372 Percent Identity: 68.992

alignment\_block:

US-09-327-750D-30 x AF237783

Align seg 1/1 to: AF237783 from: 1 to: 791

```

1 MetGluSerLysAspGln...GlyValLysAsnLeuAsnMetGluAsnAs 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 ATGGAGTCCAAAGAGAAACAGTAAACAGTCTCAGCATGGAATATGC 221

16 pHISGlnLysLysGluGluLysGluLysProGlnAspThrIleArgA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 CAACCAAGAAATGAAGAAAGAG.....CAAGTTGCTTAATAAAG 262

33 rgGluProAlaValAlaLeuLysSerGluAlaGlyLysAsnCysAlaPro 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 GGGAGGCC...TTGGCCCTCCCTTTGGATGCTGGTGAATACTGTGTGCCT 309

50 ArgGlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 AGAGGAATCGTAGCGGTTCCGCGTTAGCAGCCCATCTGCAGTATAG 359

66 gTrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluG 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 ATGGGATATGATGCATAGGCTTGGAGAACCCACAGGCAAGGATGAGAGAAG 409

83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 AGAATATGGAAGGATGGGAGGAGGTGAGACAGCTGATGGAAAGCTG 459

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspPro 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 AGGGAAGCAGTTGAGTCATAGTCTCGCGGAGTCAGCAGCTGACCCCC 509

116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 TCACCATGACCATCATGATGAGTTTGGCTTATGCC 546

```

seq\_name: gb\_pr:AF251053

seq\_documentation\_block:

LOCUS AF251053 744 bp mRNA PRI 15-APR-2001  
DEFINITION Homo sapiens X-linked protein mRNA, complete cds.

ACCESSION AF251053

VERSION AF251053.1 GI:13625167

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 744)

AUTHORS Mao,Y., Xie,Y., Zhou,Z., Zhao,W., Zhao,S., Wang,W., Huang,Y.,  
Wang,S., Tang,R., Chen,X. and Wu,C.

TITLE Direct Submission

JOURNAL Submitted (29-MAR-2000) Institute of Genetics, School of Life  
Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R.  
China

FEATURES Location/Qualifiers

source

1..744

/organism="Homo sapiens"

/db\_xref="taxon:9606"

101..487

/codon\_start=1

/product="X-linked protein"

/protein\_id="AAK34943.1"

/db\_xref="GI:13625168"

/translation="MESKEERLNNLIVENQENDEKDEQVANKGEPLALPLNVS"



```

/chromosome="x"
/map="x"
/clone="GHC-870H8"
/clone_lib="SCcv"
475..701
/note="L1 element fragment"
repeat_region
802..988
/note="L1 element fragment"
repeat_region
misc_feature
2533..4138
/note="match: multiple ESTs"
repeat_region
4189..4246
/note="29 copies of 2 mer 91 % conserved"
7496..7874
/note="L1 element fragment"
repeat_region
8022..8108
/note="L1 element fragment"
repeat_region
8302..8493
/note="L1 element fragment"
repeat_region
8585..8671
/note="L1 element fragment"
repeat_region
8769..9010
/note="L1 element fragment"
repeat_region
9259..9384
/note="L1 element fragment"
repeat_region
9624..9731
/note="MLTIC element fragment"
repeat_region
9781..9948
/note="MLTIC element fragment"
repeat_region
9890..9948
/note="MLTIB element fragment"
repeat_region
10418..10592
/partial
/note="Alu repeat: matches 298..108 of consensus"
10593..10688
/partial
/note="Alu repeat: matches 96..1 of consensus"
10698..10781
/note="L1 element fragment"
repeat_region
11112..11403
/note="Alu repeat: matches 1..308 of consensus"
11922..12004
/note="L1 element fragment"
repeat_region
12062..12308
/note="L1 element fragment"
repeat_region
12423..12707
/partial
/note="Alu repeat: matches 302..1 of consensus"
12710..12916
/note="L1 element fragment"
repeat_region
12979..13081
/note="L1 element fragment"
repeat_region
13473..13532
/note="MLT2A1 element fragment"
repeat_region
15871..15928
/note="L1 element fragment"
repeat_region
18180..18417
/note="L1 element fragment"
repeat_region
18778..19056
/note="Alu repeat: matches 1..308 of consensus"
22279..22350
/note="L1 element fragment"
repeat_region
22449..22995
/note="L1 element fragment"
repeat_region
23239..23374
/note="L1 element fragment"
repeat_region
23371..23606
/note="MLTIB element fragment"
repeat_region
23395..23541
/note="MLTIA element fragment"
repeat_region
23705..23778
/note="MLTIA element fragment"
repeat_region
23805..23960
/note="L1 element fragment"

repeat_region 24011..24541
/note="L1 element fragment"
repeat_region 24638..24770
/partial
/note="Alu repeat: matches 150..1 of consensus"
repeat_region 24829..30975
/note="L1 element fragment"
BASE COUNT 9462 a 6147 c 6129 g 9583 t
ORIGIN
alignment_scores:
Quality: 450.00 Length: 129
Ratio: 3.879 Gaps: 2
Percent Similarity: 89.922 Percent Identity: 66.667
alignment_block:
US-09-327-750D-30 x HSV870H8 ..
Align seg 1/1 to: HSV870H8 from: 1 to: 31321
1 MetGluSerLysAspGln...GlyValLysAsnLeuAsnMetGluAsnAs 16
|||||
3508 ATGGAGTCCAAAGAGGAGCAAGCGCTTAAACATCTCATCGTGGAAAATGT 3557
16 pHisGlnLysLysGluGluLysProGlnAspThrIleArgA 33
|||||
3558 CAACACAGGAAATGATGAAAAGATGAAAAGAGGAGCAAGTTGCTAATAAAG 3607
33 rGluProAlaValAlaLeuIleSerGluAlaGlyLysAsnCysAlaPro 49
|||||
3608 GGGAGGCC...TTGGCCCTACCTTTGAATGTTAGTGAATACTGTGTGCCT 3654
50 ArgGlyClyArgArgArgPheArgValArgGlnProIleAlaHisTyrAr 66
|||||
3655 AGAGGAAACCGTAGCGGTTCCCGCTTAGCGAGCCCATCTGCAGTATAG 3704
66 gTTPAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGluG 83
|||||
3705 ATGGGACATATCATAGAGCTTTGGAGAGCCACAGGCAAGGATGAGAGG 3754
83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99
|||||
3755 AGAATATGGAAGGATTGGGGAGGAGGTGAGACAGCTGATGGAAGAGCTG 3804
100 ArgGluArgGlnLeuSerHisSerHisLeuArgAlaValSerThrAspPro 116
|||||
3805 AGGGAAGAGCAGTTGAGTCATAGTCTCGGGCAGTCAGCAGCTGATCCCC 3854
116 oHisHisAspHisHisAspGluPheCysLeuMetPro 128
|||||
3855 TCACCATGACCATCAGCATGAGTTTTCCTTATGCC 3891
seq_name: gb_pr:AL133348
seq_documentation_block:
LOCUS AL133348 40584 bp DNA PRI 06-MAR-2000
DEFINITION Human DNA sequence from clone RPL-79P11 on chromosome Xq21.32-22.1.
Contains the gene for a novel protein similar to mouse Bex2
(brain-expressed X-linked protein 2), ESTs, STSs, GSSs and a
putative CpG island, complete sequence.
ACCESSION AL133348
VERSION AL133348.8 GI:7076386
KEYWORDS HTG; Bex2; CpG island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 40584)
AUTHORS Wilson,S.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

```

## COMMENT

requests: clonerequest@sanger.ac.uk  
 On Feb 24, 2000 this sequence version replaced gi:6997869.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
 on the WORMPEP database can be found at  
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/ChrX>  
 RP1-79P11 is from the library RP1-1 constructed at the Roswell  
 Park Cancer Institute by the group of Pieter de Jong. For further  
 details see <http://bacpac.med.buffalo.edu/>  
 VECTOR: pCYPAC2  
 IMPORTANT: This sequence is not the entire insert of clone  
 RP1-79P11 it may be shorter because we only sequence overlapping  
 sections once, or longer because we arrange for a small overlap  
 between neighbouring submissions.  
 The true left end of clone LLOXNC01-105G4 is at 40485 in this  
 sequence. The true right end of clone LLOXNC01-177E8 is at 100 in  
 this sequence.

## FEATURES

Source	Location/Qualifiers
repeat_region	1..40584
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="X"
repeat_region	/map="q21.32-22.1"
repeat_region	/clone="RP1-79P11"
repeat_region	/clone_lib="RP1-1"
repeat_region	19..292
repeat_region	/note="AluJo repeat: matches 3..275 of consensus"
repeat_region	470..1284
repeat_region	/note="LIMB5 repeat: matches 5318..6157 of consensus"
repeat_region	1295..1898
repeat_region	/note="L1PA15 repeat: matches 5556..6157 of consensus"
repeat_region	1899..2345
repeat_region	/note="L1PA15 repeat: matches 4922..5370 of consensus"
repeat_region	2323..2698
repeat_region	/note="LIMB5 repeat: matches 4960..5343 of consensus"
repeat_region	2927..3065
repeat_region	/note="MER74A repeat: matches 164..305 of consensus"
repeat_region	3655..9795
repeat_region	/note="L1PA4 repeat: matches 2..6142 of consensus"
repeat_region	9866..9982
repeat_region	/note="FLAW_C repeat: matches 6..124 of consensus"
repeat_region	9989..10254
repeat_region	/note="L1 repeat: matches 4027..4293 of consensus"
repeat_region	10253..10852
repeat_region	/note="L1M4 repeat: matches 4755..5333 of consensus"
repeat_region	10853..11260
repeat_region	/note="MLT1B repeat: matches 3..390 of consensus"
repeat_region	11261..11474
repeat_region	/note="L1M4 repeat: matches 4538..4755 of consensus"
repeat_region	11480..11563
repeat_region	/note="MSTB repeat: matches 1..80 of consensus"
repeat_region	11564..12475
repeat_region	/note="L1 repeat: matches 3039..3971 of consensus"
misc_feature	14463..14861
repeat_region	/note="match: GSS: Em:AQ729215"
repeat_region	15578..15854
misc_feature	16197..16569
repeat_region	/note="match: GSS: Em:AQ772810"
repeat_region	16197..16569
repeat_region	/note="L1M4 repeat: matches 4267..4627 of consensus"
repeat_region	16629..16714
repeat_region	/note="L1M4C repeat: matches 1578..1666 of consensus"
repeat_region	16902..17133
repeat_region	/note="L1M4 repeat: matches 0..242 of consensus"
repeat_region	17536..17728
repeat_region	/note="MIR repeat: matches 1..191 of consensus"
repeat_region	17746..18323
repeat_region	/note="L2 repeat: matches 1599..2175 of consensus"
repeat_region	18706..18755
repeat_region	/note="L1MB3 repeat: matches 6034..6083 of consensus"
repeat_region	18756..19236
repeat_region	/note="MER66B repeat: matches 1..486 of consensus"
repeat_region	19237..21108
repeat_region	/note="MER66-internal repeat: matches 4808..6676 of consensus"
repeat_region	21109..21163
repeat_region	/note="MLT2B repeat: matches 391..445 of consensus"
repeat_region	21164..21555
repeat_region	/note="MER66B repeat: matches 3..391 of consensus"
repeat_region	21557..21662
repeat_region	/note="L1MB3 repeat: matches 6080..6185 of consensus"
repeat_region	21665..21761
repeat_region	/note="L1M repeat: matches 5299..5392 of consensus"
repeat_region	21762..21929
repeat_region	/note="L1MB8 repeat: matches 5902..6063 of consensus"
repeat_region	21930..22212
repeat_region	/note="AluSg repeat: matches 3..284 of consensus"
repeat_region	22213..22321
repeat_region	/note="L1MB8 repeat: matches 6063..6173 of consensus"
repeat_region	22333..22740
repeat_region	/note="L1MD2 repeat: matches 5867..6288 of consensus"
repeat_region	22744..23094
repeat_region	/note="L1MD repeat: matches 2..342 of consensus"
repeat_region	23218..23524
repeat_region	/note="AluSg repeat: matches 1..305 of consensus"
repeat_region	23603..23947
repeat_region	/note="L1M4 repeat: matches 4916..5184 of consensus"
repeat_region	23948..24218
repeat_region	/note="AluJo repeat: matches 1..282 of consensus"
repeat_region	24219..24510
repeat_region	/note="L1M4 repeat: matches 4601..4916 of consensus"
repeat_region	24508..24621
repeat_region	/note="L1M4 repeat: matches 4374..4488 of consensus"
repeat_region	24650..25077
repeat_region	/note="MLT1C repeat: matches 1..466 of consensus"
repeat_region	25078..25914
repeat_region	/note="L1M4 repeat: matches 3497..4346 of consensus"
repeat_region	25920..26741
repeat_region	/note="L1M4 repeat: matches 2202..3065 of consensus"
repeat_region	26762..26906
repeat_region	/note="L1PA7 repeat: matches 6001..6145 of consensus"
repeat_region	26907..27135
repeat_region	/note="L1PA5 repeat: matches 5674..5902 of consensus"
repeat_region	27266..27509
repeat_region	/note="L1M4 repeat: matches 2164..2080 of consensus"
repeat_region	27557..27769
repeat_region	/note="L1MEC repeat: matches 253..460 of consensus"
repeat_region	28120..28167
repeat_region	/note="L2 repeat: matches 2703..2750 of consensus"
repeat_region	29043..29349
repeat_region	/note="MERS8B repeat: matches 1..341 of consensus"
repeat_region	30389..30452
repeat_region	/note="16 copies 4 mer tata 89% conserved"
polyA_site	complement(30502)
misc_feature	/gene="dJ79P11.1"
misc_feature	complement(join(30502..31137,31514..31589,32016..32134))
misc_feature	/gene="dJ79P11.1"
misc_feature	/note="match: ESTs: Em:AA317587 Em:AA358632 Em:AI214048

```
Em:AI929470 Em:AA319216 Em:AA317625 Em:DB1392 Em:AA317418
Em:AA356528 Em:C15186 Em:AI291270 Em:AA160660 Em:AA612605
Em:AI160832 Em:AA384405 Em:AA160657 Em:AA081744
Em:AI291126 Em:C15591 Em:AA384268 Em:AI929703 Em:AI302026
Em:W19547 Em:AI141727 Em:AA946933 Em:AI141583 Em:H61106
Em:W60581 Em:AI826617 Em:AA706545 Em:AA464771 Em:AA862327
Em:W68380 Em:AA641652 Em:AA384269 Em:AA527696 Em:W48778
Em:AA633989 Em:AI659486 Em:AW167301 Em:AI929428
Em:AA379411 Em:C15948 Em:AI745669 Em:AW024587 Em:AW379207;
match: proteins: Tr:Q9Y516 Tr:Q9Y517
/evidence=not_experimental
/product="dj79p11.1 (novel protein similar to mouse Bex2
complement(30502, .32134)
/gene="dj79p11.1"
30507. .30988
/note="match: STS: Em:G24641; match: STS: Em:T86927"
30508. .30852
/note="match: STS: Em:R43117; match: STS: Em:G23964"
complement(30518, .30523)
/gene="dj79p11.1"
30541. .30922
/note="match: STS: Em:H61107"
31543. .32262
/note="CpG island"
/evidence=not_experimental
31647. .31730
/note="42 copies 2 mer cc 64% conserved"
complement(32655, .33316)
/note="match: GSS: Em:AA0489303"
33027. .34514
/note="L1ME2 repeat: matches 4631. .6155 of consensus"
34528. .34786
/note="L1ME3A repeat: matches 4404. .4664 of consensus"
34787. .35085
/note="Alus9 repeat: matches 1. .300 of consensus"
35086. .35412

alignment_scores:
  Quality: 450.00      Length: 129
  Ratio: 3.879        Gaps: 2
Percent Similarity: 89.922 Percent Identity: 66.667

alignment_block:
US-09-327-750D-30 x AL133348/rev ..
Align seg 1/1 to reverse of: AL133348 from: 1 to: 40584

1 MetGluSerLysAspGln...GlyValLysAsnLeuAsnMetGluAsnAs 16
|||||
31132 ATGGAGTCCAAAGAGGAGGCGGTTAAACAATCTCATCGTGAATAATGT 31083

16 pHisGlnLysLysGluGluLysGluLysProGlnAspThrIleArgA 33
|||||
31082 CAACCCAGGAAATGATGAAAGAGATGAAAGGACGAGTGTCTAATAAAG 31033

33 rgGluProAlaValAlaLeuSerGluAlaGlyLysAsnCysAlaPro 49
|||||
31032 GGGAGGCC...TTGGCCCTACCTTTGAATGTTAGTGAATACGTGTGCCT 30986

50 ArgGlyGlyArgArgPheArgValArgGlnProIleAlaHisTyrAr 66
|||||
30985 AGAGGAAACCGTAGGCGGTTCCGCGTTAGCAGCCCATCTCGCAGTATAG 30936

66 gTrpAspLeuMetGlnArgValGlyGluProGlnGlyArgMetArgGlu 83
|||||
30935 ATGGGACATAATCATAGCTTGGAGAGCCACAGGCAAGGATGAGAGAGG 30886

83 luAsnValGlnArgPheGlyGlyAspValArgGlnLeuMetGluLysLeu 99
|||||
30885 AGAATATGGNAAGGATTTGGGAGAGGTTGAGACAGCTGATGAGAAAGCTG 30836

100 ArgGluArgGlnLeuSerHisSerLeuArgAlaValSerThrAspProPr 116
```

```
|||||:::|||||
30835 AGGAAAGCAGTTGAGTCATAGTTTTCGGGCGAGTCAGCACTGATCCCC 30786

116 oHisHisAspHisHisAspCluPheCysLeuMetPro 128
|||||
30785 TCACCATGACCATCAGCATGAGTTTTCGCCCTATGCCCC 30749

seq_name: gb_htg:HSU80B1

seq_documentation_block:
LOCUS HSU80B1 41029 bp DNA HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome X clone LLOXNC01-80B1 map q22.1-22.3, ***
SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AL022169
VERSION AL022169.3 GI:13276704
KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 41029)
AUTHORS Bird, C.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 12, 2001 this sequence version replaced gi:12750927.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: cu80B1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ABI; 19% of reads
Chemistry: Dye-terminator Big Dye; 80% of reads
Consensus quality: 41009 bases at least Q40
Consensus quality: 41023 bases at least Q30
Consensus quality: 41027 bases at least Q20
Insert size: 41029; sum-of-contigs
Insert size: 51751; 0.6% error; agarose-fp
Quality coverage: 9.67x in Q20 bases; sum-of-contigs Quality
coverage: 7.67x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
      source
      Location/Qualifiers
          1..41029
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="X"
             /map="q22.1-22.3"
             /clone="LLOXNC01-80B1"
             /clone_lib="LLOXNC01"
          1..41029
             misc_feature
             /note="assembly_fragment:00890"

BASE COUNT 13249 a 7796 c 7781 g 12203 t
ORIGIN

alignment_scores:
  Quality: 450.00      Length: 129
  Ratio: 3.879        Gaps: 2
Percent Similarity: 89.922 Percent Identity: 66.667

alignment_block:
US-09-327-750D-30 x HSU80B1
```



OM of: US-09-327-750D-34 to: EST:\* out\_format : pfs  
Date: Mar 11, 2002 2:16 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=framed\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09327750/runat\_11032002\_101153\_20308/app\_query.fasta\_1.1472  
-DB=EST -OPMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTPWT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09327750\_@CGNI\_1\_5654  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-327-750D-34  
Query length: 118  
Database: EST:  
Database sequences: 11351937  
Database length: 1077921985  
Search time (sec): 4085.940000

score\_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
gb_est1:AI227867	-	620.00	1031.44	2.7e-48	568	AI227867 EST224562 Normalized r
gb_est2:BI040889	-	592.00	984.02	1.2e-45	643	BI289546 UI-R-DKO-cff-c-12-0-UI
gb_hic:AK010400	+	546.00	906.12	2.6e-41	789	AK010400 Mus musculus ES cells
gb_est2:BG083261	+	546.00	906.10	2.6e-41	792	BG083261 H3086C08-5 NIA Mouse 1
gb_est1:AA473525	-	534.00	885.85	4.1e-39	458	AA473525 vg78a09.r1 Barstead MP
gb_est1:AW251884	-	520.00	865.45	4.8e-39	540	AW251884 UI-R-BJ0-adn-b-06-0-UI
gb_est2:BG070341	+	510.00	845.61	5.4e-38	747	BG070341 H3086C08-3 NIA Mouse 1
gb_est2:BF723075	+	503.00	837.44	1.8e-37	519	BF723075 mab27e12.y1 Soares_NME
gb_est1:BE333774	+	483.00	804.95	1.1e-35	463	BE333774 us27h06.y1 Soares_NME
gb_est2:W17712	+	466.00	779.24	3.1e-34	317	W17712 mb77e04.r1 Soares mouse
gb_est1:AA117429	+	444.00	742.75	3.3e-32	312	AA117429 mb21e12.r1 Bedington
gb_est1:AI012535	-	426.00	710.42	2.1e-30	444	AI012535 EST206986 Normalized r
gb_est2:W35893	+	425.00	708.34	2.7e-30	472	W35893 mc53g09.r1 Soares mouse
gb_est2:W54487	+	410.00	683.49	6.6e-29	465	W54487 md09c11.r1 Soares mouse
gb_est1:AL118340	+	407.00	679.16	1.2e-28	422	AL118340 v911b26 Bedington mc
gb_est2:BF722027	-	406.00	680.14	1.0e-28	286	BF722027 mab27e12.x1 Soares_NME
gb_est1:AA104150	+	399.00	665.68	6.9e-28	433	AA104150 mc50h03.r1 Life Tech m
gb_est2:W85403	+	388.00	647.24	6.9e-27	442	W85403 mf58d02.r1 Soares mouse
gb_est1:AA893925	-	386.00	644.71	9.5e-27	393	AA893925 EST197728 Normalized r
gb_est1:AA956955	-	385.00	641.88	1.4e-26	467	AA956955 UI-R-E1-fl-f-07-0-UI.S
gb_est2:BF556502	+	373.00	623.91	1.4e-25	348	BF556502 UI-R-E1-fl-f-07-0-UI.r
gb_est1:AL517301	+	372.00	618.13	2.9e-25	639	AL517301 AL517301 LTI_NFL011.NE
gb_est2:BE883414	+	372.00	617.30	3.2e-25	722	BE883414 601509167F1 NIH.MGC.71
gb_est1:AL544124	+	372.00	616.23	3.7e-25	845	AL544124 AL544124 LTI_NFL006.PI
gb_est1:AUI17899	+	372.00	616.08	3.8e-25	865	AUI17899 HEMBA1 Homo S
gb_est2:BG122197	+	372.00	616.07	3.8e-25	866	BG122197 602347948F1 NIH.MGC.90
gb_est1:AL545120	+	372.00	615.77	3.9e-25	905	AL545120 AL545120 LTI_NFL006.PI
gb_est2:BG09427	+	372.00	615.59	4.0e-25	929	BG09427 602673481F1 NIH.MGC.96
gb_est2:BG035675	+	372.00	615.26	4.2e-25	975	BG035675 602325745F1 NIH.MGC.90
gb_est1:AL560650	+	372.00	615.21	4.2e-25	983	AL560650 AL560650 LTI_NFL010.BC
gb_est1:AL539004	+	372.00	614.94	4.3e-25	1023	AL539004 AL539004 LTI_FL013.FB
gb_est1:AL549875	+	372.00	614.80	4.4e-25	1044	AL549875 AL549875 LTI_NFL006.F
gb_est2:BF930215	+	367.50	611.31	6.9e-25	579	BF930215 IL5-NT0277-111200-319-
gb_est1:AL534630	+	367.00	610.64	7.5e-25	566	AL534630 AL534630 LTI_FL013.FB
gb_est1:AL523320	+	365.50	603.99	1.8e-24	1043	AL523320 AL523320 LTI_NFL003.N
gb_est1:AW385267	+	364.00	606.04	1.4e-24	534	AW385267 RC0-LT0001-261199-011-
gb_est2:BG527431	+	364.00	603.54	1.9e-24	772	BG527431 602557227F1 NIH.MGC.59
gb_est2:BG009563	-	362.50	602.97	2.0e-24	581	BG009563 QV1-GN0319-021200-526-
gb_est1:AL573904	+	362.00	597.32	2.7e-24	1182	AL573904 AL573904 LTI_NFL006.F
gb_est2:BI040887	-	360.50	600.82	2.7e-24	489	BI040887 QV3-NT0278-120201-502-

gb\_est1:AL582060 - 359.50 594.05 6.3e-24 1037 ! AL582060 AL582060 LTI\_NFL01  
gb\_est2:BI040889 + 356.50 593.18 7.1e-24 565 ! BI040889 QV3-NT0278-120201-5  
gb\_est1:AW934883 + 355.00 589.55 1.1e-23 668 ! AW934883 RC1-DT0001-230200-0  
gb\_est1:AW934905 + 355.00 589.38 1.2e-23 685 ! AW934905 RC1-DT0001-230200-0  
gb\_est1:AW934769 + 354.50 588.78 1.2e-23 662 ! AW934769 RC1-DT0001-230200-0  
seq\_name: gb\_est1:AI227867  
seq\_documentation\_block:  
LOCUS AI227867 568 bp mRNA EST 20-JAN-1999  
DEFINITION EST224562 Normalized rat brain, Bento Soares Rattus sp. cDNA clone  
RBCM82 3' end, mRNA sequence.  
ACCESSION AI227867  
VERSION AI227867.1 GI:3811754  
KEYWORDS EST.  
SOURCE Rattus sp.  
ORGANISM Rattus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 568)  
AUTHORS Lee N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,  
Kerlavager, A.R. and Adams, M.D.  
TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat  
Gene Index  
JOURNAL Unpublished (1998)  
COMMENT Other ESTs: TC58086  
Contact: Lee, NH  
The Institute for Genomic Research  
9712, Medical Center Drive, Rockville, MD 20850, USA  
Tel: (301)-838-3529  
Fax: (301)-838-0208  
Email: nhlee@tigr.org  
Seq primer: M13-21.  
Location/Qualifiers  
source 1..568  
/organism="Rattus sp."  
/db\_xref="ATCC (inhost):2036891"  
/db\_xref="taxon:10118"  
/clone="RBCM82"  
/clone\_lib="Normalized rat brain, Bento Soares"  
/note="Organ: brain; Vector: pT73pac; Site\_1: EcoRI;  
Site\_2: NotI"  
BASE COUNT 140 a 143 c 108 g 177 t  
ORIGIN  
alignment\_scores:  
Quality: 620.00 Length: 118  
Ratio: 5.299 Gaps: 0  
Percent Similarity: 99.153 Percent Identity: 99.153  
alignment\_block:  
US-09-327-750D-34 x AI227867/rev ..  
Align seg 1/1 to reverse of: AI227867 from: 1 to: 568  
1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17  
|||||  
528 ATGGCGTCCAAAGTCAAAAGTCATCTGGATCTCCTGCGGAGAAAGA 479  
|||||  
17 pLysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34  
|||||  
478 CAAAGAAAACAAAAAGTGGGAGGCGCTCCAAACAAAGTGAAGAAAT 429  
|||||  
34 erHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50  
|||||  
428 CCACCATCTGGAAGAGGTGTAACAAACAAAGAGCTGGGGCAATGTCAG 379  
|||||  
51 ArgLysValArgArgLeuValProAsnPhleuTrpAlaIleProAsnAr 67  
|||||  
378 AGAAAGTCAGGCGACTTGTGCCTAACTTCTATGGCCATACCTAATAG 329  
|||||

67 gHisValAspHisSerGluGlyGlyGluGluValGlyArgPheValGlyG 84  
|||||  
328 GCATGTTGATCACAGTGAAGGGGAGAGGTTGGAGATTCGTAGGCG 279  
|||||  
84 lnValMetGluAlaLysArgHisSerLysGluGlnMetArgProTyr 100  
|||||  
278 AGGTGATGGGAAGCCAGAGAAAGTCTAAGGAGCAACAGATGAGGCCCTTAC 229  
|||||  
101 ThrArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeuI 117  
|||||  
228 ACCGGTTTCCGAACCCCTGAACCTGACAATATTACGACTTTTGCCCTCAT 179  
|||||  
117 ePro 118  
|||||  
178 ACCT 175

seq\_name: gb\_est2:BI289546

seq\_documentation\_block:  
LOCUS BI289546 643 bp mRNA EST 19-JUL-2001  
DEFINITION UI-R-DK0-cff-c-12-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone  
UI-R-DK0-cff-c-12-0-UI 3', mRNA sequence.  
ACCESSION BI289546  
VERSION BI289546.1 GI:14947228  
KEYWORDS EST.  
SOURCE Norway rat.  
Rattus norvegicus  
ORGANISM

REFERENCE 1 (bases 1 to 643)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
normalized rat kidney pool library cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com) The following repetitive  
elements were found in this cDNA sequence: 1-51,  
>AT\_rich#Low\_complexity  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES  
Source Location/Qualifiers  
1..643  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DK0-cff-c-12-0-UI"  
/clone\_lib="UI-R-DK0"  
/dev\_stage="ADULT"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-DK0  
library is a subtracted library derived from a mixture of  
five individually tagged normalized rat libraries:  
brain-nRBP (20%), heart-nRBP (20%), kidney-nRBP (20%),  
aorta-nRAP (20%), and placenta-nRBP (20%). Each original  
library was constructed from a mixture of equal amounts of  
RNA from seven different developmental time-points:  
embryonic day 17, embryonic day 19, embryonic day 21,

adult day 1, adult day 12, adult day 75, and adult day  
200. (Exception: the aorta pool does not contain embryonic  
day 17 RNA and the placenta pool contains only the three  
embryonic stages). Each library was normalized  
individually according to the procedure described by  
Bonaldo, Lennon & Soares (Genome Research Genome 6:  
791-806, 1996). For construction of the DK0 subtracted  
library, plasmid DNA from each of the five individually  
tagged normalized libraries was mixed in the proportions  
specified above and electroporated into competent bacteria  
for production of single-stranded circular DNA  
representing the pool of libraries. Single-stranded  
circular DNA representing these five normalized libraries  
was then used as a tracer in a subtractive hybridization  
with a driver (PCR amplified inserts from a plasmid DNA  
template preparation) comprising: a) a set of about 1,000  
arrayed clones from each of the five non-normalized  
libraries of brain (CT0s), heart (CS0s), kidney (CU0s),  
aorta (CW0s), and placenta (CX0s). The resulting pool of  
approximately 5,000 clones represented about 33.3% of the  
final driver population. A set of about 2,000 arrayed  
clones from each of the five normalized libraries of brain  
(CT0), heart (CS0), kidney (CU0), aorta (CW0), and  
placenta (CX0). The resulting pool of about 10,000 clones  
represented about 66.6% of the final driver population.

TAG-LIB=UI-R-DK0  
TAG-TISSUE=rat kidney pool  
TAG\_SEQ=CAAGACTGTCTC  
BASE COUNT 160 a 149 c 113 g 221 t  
ORIGIN

alignment\_scores:  
Quality: 592.00 Length: 118  
Ratio: 5.060 Gaps: 0  
Percent Similarity: 99.153 Percent Identity: 98.305

alignment\_block:  
US-09-327-750D-34 x BI289546/rev ..

Align seg 1/1 to reverse of: BI289546 from: 1 to: 643

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17  
601 ATGCGGTCCAAAGTCAAAACAAAGTCATCTGATCTCCTCTGGGAGAAAGA 552  
17 pLysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34  
551 CAAGAAAACAAAAAGGTGGGAG.GCCTCCAAACAAAGTGAAGAGAAT 503  
34 erHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50  
502 CCCACCATCTGGGAAGAGGTTGAAACAAAGAGCCTGGGGCAATGTCCAGG 453  
51 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 67  
452 AGGAAAGTCAGGCGACTTGTGCTTAACCTTTCTATGGGCCCATCTAATAG 403  
67 gHisValAspHisSerGluGlyGlyGluGluValGlyArgPheValGlyG 84  
402 GCATGTTGATCACAGTGAAGGGGAGAGGTTGGGAGATTCGTAGGCGC 353  
84 lnValMetGluAlaLysArgHisSerLysGluGlnMetArgProTyr 100  
352 AGGTGATGGAAGCCAGAGAAAGTCTAAGGAGCAACAGATGAGGCCCTTAC 303  
101 ThrArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeuI 117  
302 ACCGGTTTCCGAACCCCTGAACCTGACAATATTACGACTTTTGCCCTCAT 253  
117 ePro 118  
|||||  
252 ACCT 249



```

seq_name: gb_hlc:AK010400
seq_documentation_block:
LOCUS AK010400 789 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library,
clone:2410004M13, full insert sequence.
ACCESSION AK010400
VERSION AK010400.1 GI:12845816
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA.
clone_lib:RIKEN full-length enriched mouse cDNA library
clone:2410004M13.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 789)
Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
2 (bases 1 to 789)
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalizaton and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED 11042159
3 (bases 1 to 789)
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,K., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipillar sequencer
Genome research. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED 11076861
4 (bases 1 to 789)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED
REFERENCE
5 (bases 1 to 789)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
TITLE Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in RIKEN
Genomic Sciences Center and Genome Science Laboratory in RIKEN.

```

Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5' GAGAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. cDNA went  
through one round of normalization to Rot = 5.0 and subtraction to  
Rot = 25.0. Second strand cDNA was prepared with the primer adapter  
of sequence [5'  
GAGAGAGAGATTCGAGTTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved  
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.

Host: SOLR.

FEATURES  
Location/Qualifiers  
1..789  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGI:1328321"  
/db\_xref="MGI:1908964"  
/clone="2410004M13"  
/cell\_type="ES cells"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
208..564  
/note="putative"  
/codon\_start=1  
/protein\_id="BAB26911.1"  
/db\_xref="GI:12845817"  
/translation="MASRFKOVILDLTVEKDKDKKGGKASKQSEEPHLEEVENKK  
PGNNVRKVRBLVNFNLWALPNRHRVDRNEGDEGVRVVQGVTEVKKRKTTEQQVRP  
FRTPEDNHYDFCLIP"  
770..775  
/note="putative"  
789  
/note="putative"  
polyA\_signal 239 a 161 c 205 g 184 t  
polyA\_site  
BASE COUNT  
ORIGIN  
alignment\_scores  
Quality: 546.00 Length: 118  
Ratio: 4.964 Gaps: 0  
Percent Similarity: 93.220 Percent Identity: 87.288  
alignment\_block  
US-09-327-750D-34 x AK010400 ..  
Align seg 1/1 to: AK010400 from: 1 to: 789  
1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17  
||||| 17  
208 ATGGCATCCAAATTTAAACAAGTCATCTGATCTCACTGTGGAGAAAGA 257  
17 pLysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34  
||||| 34  
258 CAAAAGAACAAAAAGGTGGGAGGCTCCAAACAAAGTGAAGAAGAAC 307  
34 erHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50  
||||| 50  
308 CCACCATCTGGAGAGGTGGAACACAGAGGCTGGGGAAATGTCCGA 357  
51 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 67  
||||| 67  
358 AGGAAAGTCAGGCGACTGTGCGCTAACTTCTCTGGGCCATACCAATAG 407  
67 gHisValAspHisSerGlyGlyGluGluValGlyArgPheValGlyG 84  
||||| 84  
408 GCATGTTGATCGCAATCAAGGGGAGAGGATGTTGGAGATTTGTAGTGC 457  
84 lnValMetGluAlaLysArgHisSerLysGluGlnGlnMetArgProTyr 100  
||||| 100  
458 AGGACACAGAGTCAGAGAGAAAGACTACGAGGAGGAGGTCAGGCTTAC 507  
101 ThrArgPheArgThrProGluProAsnHisTyrAspPheCysLeu 117  
||||| 117

508 AGCGTTTCGACACCCGGAACCTGACAATCATTACGACTTTTGCCTCAT 557

117 ePro 118

|||||  
558 ACCT 561

seq\_name: gb\_est2:BG083261

seq\_documentation\_block: 792 bp mRNA EST 26-JAN-2001  
LOCUS BG083261  
DEFINITION H3086C08-5 NIA Mouse 15K CDNA Clone Set Mus musculus CDNA clone  
H3086C08 5', mRNA sequence.  
ACCESSION BG083261  
VERSION BG083261.1 GI:12565829  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 792)  
AUTHORS Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka  
,T.S., Carter,M.G. and Ko,M.S.H.  
TITL Verification and initial annotation of NIA mouse 15K CDNA clone set  
JOURNA Unpublished (2001)  
COMMENT Other ESTs: H3086C08-3

Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@lgsun.grc.nia.nih.gov  
This clone set has been freely distributed to the community. Please  
visit <http://lgsun.grc.nia.nih.gov/cdna/15k.html> for details.  
Plate: H3086 row: C column: 08  
Seq primer: -21M13 Reverse  
High quality sequence stop: 792  
POLYA-No.

#### FEATURES

Location/Qualifiers  
1..792  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="H3086C08"  
/clone.lib="NIA Mouse 15K CDNA Clone Set"  
/sex="Clones arrayed from a variety of CDNA libraries"  
/dev\_stage="Clones arrayed from a variety of CDNA libraries"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo CDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary CDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT) Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental CDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale CDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."  
BASE COUNT 247 a 153 c 210 g 179 t 3 others  
ORIGIN

#### alignment\_scores:

Quality: 546.00 Length: 118  
Ratio: 4.964 Gaps: 0

Percent Similarity: 93.220 Percent Identity: 87.288

#### alignment\_block:

US-09-327-750D-34 x BG083261

Align seg 1/1 to: BG083261 from: 1 to: 792

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17  
|||||  
190 ATGGCATCCAAATTTAAACAAGTCATACCTGCTCCTGTTGGAGAAAGA 239  
17 pLysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34  
|||||  
240 CAAAAAGACAAAAAGGTGGGAAGGCCTCCAAACAAGTGAAGAAGAAC 289  
34 erHisLysLeuGluValGluAsnLysLysProGlyGlyAsnValArg 50  
|||||  
290 CCCACCATCTGGAAGAGGTGTAACAACAAGAGCTGGGGAAATGTCGA 339  
51 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 67  
|||||  
340 AGAAAGTCAGCGGACTTGTGCTTAACCTTCTCTGGGCCATACCAATAG 389  
67 gHisValAspHisSerGluGlyGlyGluValGlyArgPheValGly 84  
|||||  
390 GCATGTTGATCCCAATGAAGGGGAGAGGATGTTGGAGCATTTGTAGTGC 439  
84 lnValMetGluAlaLysArgHisSerLysGluGlnGlnMetArgProTyr 100  
|||||  
440 AGGGAACAGAAAGTCAAGACAAAGACATACGAGCAGGAGGCGCTTAC 489  
101 ThrArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeu 117  
|||||  
490 AGCGGTTTCGAAACCCCGAACCTGACAATCATACGACTTTTGCCTCAT 539  
117 ePro 118  
|||||  
540 ACCT 543

seq\_name: gb\_est1:AA473525

seq\_documentation\_block:

LOCUS AA473525 458 bp mRNA EST 18-JUN-1997  
DEFINITION vg78a09.r1 Barstead MPLRBL Mus musculus. CDNA clone IMAGE:872056 5'  
similar to SW:HG74\_HUMAN Q00994 OVARIAN GRANULOSA CELL 13.0 KD  
PROTEIN HGR74. ?, mRNA sequence.

ACCESSION AA473525

VERSION AA473525.1 GI:2201752

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 458)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

MGI:511536

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 417.

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 540)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
normalized atrium at 16.5 dpc library cDNA Library Preparation:  
M.B. Soares Lab Clone Distribution: clones will be available  
through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

Location/Qualifiers  
1..540  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BJ0-adn-b-06-0-UI"  
/clone\_lib="UI-R-BJ0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-BJ0  
library is a subtracted library derived from the UI-R-BJ0  
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and  
UI-R-AG1 libraries. These libraries represent tissues f  
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV cana  
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV  
canal at 15 dpc, and ventricle at 13 dpc. The tag is a  
string of 5-6 nucleotides present between the Not I site  
and the oligo-dT track. The library was constructed as  
described by Bonaldo, Lennon and Soares, Genome Resear  
6: 791-806, 1996.  
TAG\_LIB=UI-R-BJ0  
TAG\_USAGE=atrium at 16.5 dpc  
TAG\_SEQ=GATTC"  
BASE COUNT 145 a 120 c 94 g 180 t  
ORIGIN

alignment\_scores:  
Quality: 520.00 Length: 97  
Ratio: 5.417 Gaps: 0  
Percent Similarity: 98.969 Percent Identity: 98.969

alignment\_block:  
US-09-327-750D-34 x AW251884/rev ..  
Align seg 1/1 to reverse of: AW251884 from: 1 to: 540

22 LysGlyGlyLysAlaSerLysGlnSerGluGluSerHisHisLeuGl 38  
|||||  
538 AARGTGGGAAGGGCTCCAAACAAGTGAAGAGGAATCCCACTCTTGA 489  
|||||  
38 uGluValGluAsnLysLysProGlyGlyAsnValArgArgLysValArgA 55  
|||||  
488 AGAGTTGAAACAAAGAGAGCTGGGGGCAGTCTCAGGAGGAAGTCAGGC 439  
|||||  
55 rgLeuValProAsnPheLeuTrpAlaIleProAsnArgHisValAspHis 71  
|||||

438 GACTTGTGCGCTAACTTTCTATGGCCCATACCTAATAGGCATGTTGATCAC 389  
 72 SerGluGlyGlyGluGluValGlyArgPheValGlyGlnValMetGluAl 88  
 |||||  
 388 AGTGAAGGGGAGAGAGAGGTTGGGAGATTCCTAGGCAGGATGATGGAAC 339  
 88 aLysArgHisSerLysGluGlnGlnMetArgProTyrThrArgPheArgT 105  
 |||||  
 338 CAAGAGAAAGCTCAAGGAGCAACAGATGAGGCCTTACACGCGTTCCGAA 289  
 105 hrProGluProAspAsnHisTyrAspPheCysLeuIlePro 118  
 288 CCCCTGAACCTGACAATCATTTAGGACTTTTGCCTCATACCT 248

seq\_name: gb\_est2:BG070341

seq\_documentation\_block: 747 bp mRNA EST 26-JAN-2001  
 LOCUS BG070341 H3086C08-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
 DEFINITION H3086C08 3', mRNA sequence.  
 ACCESSION BG070341  
 VERSION BG070341.1 GI:12552910  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 747)  
 Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka  
 T.S., Carter, M.G. and Ko, M.S.H.  
 Verification and initial annotation of NIA mouse 15K cDNA clone set  
 Unpublished (2001)  
 Other ESTs: H3086C08-5  
 Contact: George J. Kargul  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgsun.grc.nia.nih.gov  
 This clone set has been freely distributed to the community. Please  
 visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.  
 Plate: H3086 row: C column: 08  
 Seq primer: -21M13 Forward  
 High quality sequence stop: 747  
 POLYA=Yes.

FEATURES  
 Source

Location/Qualifiers  
 1..747  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="H3086C08"  
 /clone.lib="NIA Mouse 15K cDNA Clone Set"  
 /sex="Clones arrayed from a variety of cDNA libraries"  
 /dev\_stage="Clones arrayed from a variety of cDNA libraries"  
 /lab\_host="DH10B"  
 /note="Vector: pSPOR1; Site 1: SalI; Site 2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT) primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum

BASE COUNT 178 a 198 c 140 g 231 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 510.00 Length: 118  
 Ratio: 4.722 Gaps: 0  
 Percent Similarity: 91.525 Percent Identity: 85.593  
 alignment\_block:  
 us-09-327-750D-34 x BG070341/rev ..  
 Align seg 1/1 to reverse of: BG070341 from: 1 to: 747  
 1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17  
 |||||  
 581 ATGGCATCCAAATTTAAACAAGTCATACCTGATCTCAGTCTGGAGAAGA 532  
 |||||  
 17 pLysLysAsnLysLysGlyLysAlaSerLysGlnSerGluGluGluS 34  
 |||||  
 531 CAAAAAGACAAAAAGGTGGGAAGCCTCCAAACAAAGTGAAGAGAAC 482  
 |||||  
 34 erHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50  
 |||||  
 481 CCCACCATCTGGAAGAGGTGTAACACAAAGCCTGGGGAAATGTCCGA 432  
 |||||  
 51 ArgLysValArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 67  
 |||||  
 431 AGGAAAGTCAGGCGACTTGTGCCTAACTTTCTCTGGGCCATACCAATAG 382  
 |||||  
 67 gHisValAspHisSerGluGlyGlyGluGluValGlyArgPheValGlyG 84  
 |||||  
 381 GCATGTTGATCGCAATGAAGGGGAGAGGATTTGGGAGATTGTGTAGTGC 332  
 |||||  
 84 InValMetGluAlaLysArgHisSerLysGluGlnGlnMetArgProTyr 100  
 |||||  
 331 AGGGAACAGAGTCAGAGAGAAAGACTACGAGCAGCAGGTGACCGT.TAC 283  
 |||||  
 101 ThrArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeuI 117  
 |||||  
 282 AGGCGTTTCCGAACCCCGGAACCTGACATCATTAGGACTTTTGCCTCAT 233  
 |||||  
 117 ePro 118  
 |||||  
 232 ACCT 229

seq\_name: gb\_est2:BF723075

seq\_documentation\_block:

LOCUS BF723075 519 bp mRNA EST 03-JAN-2001  
 DEFINITION mab27e12.y1 Soares\_NMEBA\_branchial\_arch Mus musculus cDNA clone  
 IMAGE:3971447 5' similar to TR:Q9Y516 Q9Y516 D3635G19.2.3 ;, mRNA  
 sequence.  
 ACCESSION BF723075  
 VERSION BF723075.1 GI:12024077  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 519)  
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Other ESTs: mab27e12.x1  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
 , Ph.D.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml

MGI:1471479

Seq primer: -40RP from Gibco  
High quality sequence stop: 472.  
Location/Qualifiers

#### FEATURES

source

1. 519  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3971447"  
/tissue\_type="branchial\_arches"  
/dev\_stage="embryo, 10.5 dpc"  
/lab\_host="DH10B (phage resistant)"  
/note="Vector: pMT3D-Pac (Pharmacia) with a modified polylinker; Site.1: NotI; Site.2: EcoRI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAAGTGGGCGCGCATGATTTTGTGAGGAAAGA 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7p3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 158 a 109 c 170 g 81 t 1 others  
ORIGIN

#### alignment\_scores:

Quality: 503.00 Length: 113  
Ratio: 4.883 Gaps: 0  
Percent Similarity: 91.150 Percent Identity: 84.956

#### alignment\_block:

US-09-327-750D-34 x BF723075

Align seg 1/1 to: BF723075 from: 1 to: 519

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGlulLysAs 17  
181 ATGGCATCCAAATTTAAACAAGTCATACCTGGATCTCAGTGTGGAGAAAGA 230  
17 pLysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34  
231 CAAAAAAGACAAAAGGTTGGAAGGCTCCAAACAAAGTGAAGAAGAAC 280  
34 erHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50  
281 CCCACCATCTGGAAGAGGTTGAAAACAAAGAGCCTGGGGAAATGTCCGA 330  
51 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 67  
331 AGGAAGTCAGCGACTTGTGCTACTTCTCTGGGCCATACCTAATAG 380  
67 gHisValAspHisSerGluGlyGlyGluGluValGlyArgPheValGlyG 84  
381 GCATGTTGATCGCAATGAAGGGGAGAGGATGTTGGAGATTTGTAGTGC 430  
84 InValMetGluAlaLysArgHisSerLysGluGlnMetArgProTyr 100  
431 AGGGACACAGANGTCAAGAGAAAGACTACGCGACGAGGTGAGGCTTAC 480  
101 ThrArgPheArgThrProGluProAsnHisTyrAsp 113  
481 AGCGCTTCCGACGCCGGACCTGACAATCATTACGAC 519

seq\_name: gb\_est1:BE333774

seq\_documentation\_block:

LOCUS BE333774 463 bp mRNA EST 14-JUL-2000  
DEFINITION us27h06.y1 Soares\_NMEBA\_branchial\_arch Mus musculus cDNA clone  
IMAGE:3168347 5' similar to SW:HG74\_HUMAN Q00994 OVARIAN GRANULOSA

#### ACCESSION

BE333774

#### VERSION

BE333774.1

#### KEYWORDS

EST.

#### SOURCE

house mouse.

#### ORGANISM

Mus musculus

#### REFERENCE

1 (bases 1 to 463)

#### AUTHORS

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

#### TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

#### JOURNAL

Tumor Gene Index

#### COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: ccapbs-r@mail.nih.gov

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo

, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

image.llnl.gov/image/html/iresources.shtml

MGI:1063807

Seq primer: -40RP from Gibco.

Location/Qualifiers

1. 463

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:3168347"

/clone\_lib="Soares\_NMEBA\_branchial\_arch"

/tissue\_type="branchial\_arches"

/dev\_stage="embryo, 10.5 dpc"

/lab\_host="DH10B (phage resistant)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site.1: NotI; Site.2: EcoRI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5',

TGTACCAATCTGAAGTGGGCGCGCATGATTTTGTGAGGAAAGA 3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not

I and Eco RI sites of the modified pT7p3 vector. Library

constructed and normalized by Bento Soares and M.Fatima

Bonaldo."

155 a 91 c 147 g 70 t

#### BASE COUNT

ORIGIN

#### alignment\_scores:

Quality: 483.00 Length: 113  
Ratio: 4.689 Gaps: 0  
Percent Similarity: 91.150 Percent Identity: 84.956

#### alignment\_block:

US-09-327-750D-34 x BE333774

Align seg 1/1 to: BE333774 from: 1 to: 463

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGlulLysAs 17  
126 ATGGCATCCAAATTTAAACAAGTCATACCTGGATCTCAGTGTGGAGAAAGA 175  
17 pLysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34  
176 CAAAAAAGACAAAAGGTTGGAAGGCTCCAAACAAAGTGAAGAAGAAC 225  
34 erHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50  
226 CCCACCATCTGGAAGAGGTTGAAAACAAAGAGCCTGGGGAAATGTCCGA 275  
51 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 67  
276 AGGAAAGTCAGCGACTTGTGCTAACTTCTCTGGGCCATACCAATAG 325

CELL 13.0 KD PROTEIN HGR74. ;, mRNA sequence.

BE333774

BE333774.1

GI:9207550

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 463)

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: ccapbs-r@mail.nih.gov

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo

, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

image.llnl.gov/image/html/iresources.shtml

67 gHisValAspHisSerGluGlyGluValGlyArgPheValGlyG 84  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 326 GCATGTTGATCGCAATCAAGGGGAGGAGATGTTGGGAGATTTCAGTG.C 374  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 375 AGGCAACAGAGTCAAGAGAAAGACTACGGAGCAGCAGGTGAGGCCTTAC 424  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 101 ThrArgPheArgThrProGluProAspAsnHisTyrAsp 113  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 425 AAGGTTTCCGACCCCGGAACTGACCAATCATTACGAC 463

seq\_name: gb\_est2:W17712

seq\_documentation\_block: 317 bp mRNA EST 10-SEP-1996  
 LOCUS W17712 mb77e04.r1 Soares mouse p3NMf19.5 Mus musculus cDNA clone  
 DEFINITION IMAGE:335454 5' similar to SW:HG74\_HUMAN Q00994 OVARIAN GRANULOSA  
 CELL 13.0 KD PROTEIN HGR74. [1] ; mRNA sequence.

ACCESSION W17712  
 VERSION W17712.1 GI:1292113

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

REFERENCE  
 AUTHORS  
 1 (bases 1 to 317)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Scheellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennan,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The WashU-HIMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HIMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:216854

Seq primer: mob.REGA+ET

High quality sequence stop: 291.

Location/Qualifiers

FEATURES

source

1..317  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:335454"  
 /clone\_lib="Soares mouse p3NMf19.5"  
 /dev\_stage="19.5 dpc total fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Vector: pT7T3D (Pharmacia) with a modified  
 polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TCTTACCAATCTGAGTGGGAGCGCGCATTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT7T3 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M.Fatima Bonaldo. RNA was kindly provided by  
 Dr. Minoru Ko (Wayne State University)."

BASE COUNT 108 a 63 c 87 g 59 t

ORIGIN

alignment\_scores:

Quality: 466.00 Length: 101

Ratio: 5.011 Gaps: 0

Percent Similarity: 92.079 Percent Identity: 85.149

alignment\_block:

US-09-327-750D-34 x W17712

Align seg 1/1 to: W17712 from: 1 to: 317

18 LysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlucose 34

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

1 AAAAAAGACAAAAAGGTGGGAAGGCTCCAAACAAAGTGAAGAAGACC 50

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

34 rHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 51

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

51 CCACCATCTCGAAGAGGTTCAAAAACAAGAGCTGGGGAATGTCGGA 100

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

51 rGlyValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnArg 67

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

101 CGAAGTCAGCGGACCTTGTCCTTAACCTTCTCGGCCATACCAATAGG 150

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

68 HisValAspHisSerGluGlyGlyGluGluValGlyArgPheValGly 84

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

151 CATGTTGATCCCAATGAAGGGGAGAGGATGTTGGGAGATTGTAGTGCA 200

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

84 nValMetGluAlaLysArgHisSerLysGluGlnGlnMetArgProTyr 101

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

201 GCGAACAGAGAGTCAAGAGAAAGACTACGGAGCAGCAGGTGAGGCCTTACA 250

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

101 hrArgPheArgThrProGluProAspAsnHisTyrAspPheCysLeu 117

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

251 GCGGTTTCCGAAACCCGGAACCTGACAATCATTACGACTTTTGGCCTGATA 300

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

118 pro 118

|||||

301 CCT 303

seq\_name: gb\_est1:AA117429

seq\_documentation\_block:

LOCUS AA117429 312 bp mRNA EST 15-NOV-1996

DEFINITION mn2le12.r1 Beddington mouse embryonic region Mus musculus cDNA

clone IMAGE:538606 5' similar to SW:HG74\_HUMAN Q00994 OVARIAN

GRANULOSA CELL 13.0 KD PROTEIN HGR74. ; mRNA sequence.

ACCESSION AA117429

VERSION AA117429.1 GI:1672442

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 312)

AUTHORS

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Scheellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennan,G., Soares,B., Wilson,R. and

Waterston,R.

TITLE The WashU-HIMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HIMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:325542

Seq primer: -40ml3 ET

High quality sequence stop: 287.

Location/Qualifiers

1..312

FEATURES

source

/organism="Mus musculus"

/strain="C57BL6 x DBA"

/db\_xref="taxon:10090"

/clone="IMAGE:538606"  
 /clone\_lib="Beddington mouse embryonic region"  
 /sex="pooled"  
 /tissue\_type="embryo"  
 /dev\_stage="7.5dpc"  
 /lab\_host="DH12S"

/note="Organ: whole embryo; Vector: pCMV-SPORT; Site\_1:  
 SalI; Site\_2: NotI; Cloned unidirectionally. Primer:  
 Oligo dt. Gastrulating embryos were collected at 7.5dpc  
 from C57Bl/6 x DBA matings, excluding embryos that had  
 developed head folds and all extraembryonic tissues.  
 Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).  
 Referenced in Development 121, 2479-2489 (1995)"

BASE COUNT 110 a 56 c 92 g 54 t  
 ORIGIN

alignment\_scores:  
 Quality: 444.00 Length: 100  
 Ratio: 4.774 Gaps: 0  
 Percent Similarity: 93.000 Percent Identity: 86.000

alignment\_block:  
 US-09-327-750D-34 x AAL17429 ..

Align seg 1/1 to: AAL17429 from: 1 to: 312

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17  
 12 ATGGGATCCAAATTTAAACAAGTCATACCTGGATCTCCTGCGGAAAGA 61  
 17 pLysLysAsnLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34  
 62 CAAAAAGACAAAGAGGTGGGAAGGCTCCAAACAAAGTGAAGAGAAC 111  
 34 exHisHisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50  
 112 CCCACCATCTGGAAGAGGTGGAACAAAGAGCCTGGGGAAATGTCGCA 161  
 51 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 67  
 162 AGGAAAGTCAGCGCACTTGTGCTTAACCTTCTGGGCCATACCTAATAG 211  
 67 gHisValAspHisSerGluGlyGlyGluValGlyArgPheValGlyG 84  
 212 GCATGTGTCATCCATGAGGGGAGAGGATGTTGGGAGATTTGTAGTCC 261  
 84 InValMetGluAlaLysArgHisSerLysGluGlnGlnMetArgProTy 100  
 262 AGGGAACAGAGTCAAGAGAAAGACTACGAGCAGCAGGTGAGGCTTAC 311

seq\_name: gb\_est1:AI012535

seq\_documentation\_block:  
 LOCUS AI012535 444 bp mRNA EST 15-JUN-1998  
 DEFINITION EST206986 Normalized rat placenta, Bento Soares Rattus sp. cDNA  
 clone RPLAY54 3' end, mRNA sequence.

ACCESSION AI012535  
 VERSION AI012535.1 GI:3226367  
 KEYWORDS EST.  
 SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE 1 (bases 1 to 444)  
 AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,  
 Kerlavage,A.R. and Adams,M.D.

TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat

Gene Index

JOURNAL Unpublished (1998)

COMMENT Contact: Lee, NH

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208  
 Email: nhlee@tigr.org  
 Seq primer: M13-21.

FEATURES  
 source  
 1..444  
 /Organism="Rattus sp."  
 /db\_xref="taxon:10118"  
 /clone="RPLAY54"  
 /clone\_lib="Normalized rat placenta, Bento Soares"  
 /note="Organ: placenta; Vector: p773pac; Site\_1: EcoRI;  
 Site\_2: NotI"

BASE COUNT 135 a 104 c 81 g 124 t  
 ORIGIN

alignment\_scores:  
 Quality: 426.00 Length: 80  
 Ratio: 5.462 Gaps: 0  
 Percent Similarity: 97.500 Percent Identity: 97.500

alignment\_block:  
 US-09-327-750D-34 x AI012535/rev ..

Align seg 1/1 to reverse of: AI012535 from: 1 to: 444

39 GluValGluAsnLysLysProGlyGlyAsnValArgArgLysValArgAr 55  
 444 GAGGTGAAACATGAAGCCTGGGGCAATGTCAGGAGGAAAGTCAGGCG 395  
 55 gLeuValProAsnPheLeuTrpAlaIleProAsnArgHisValAspHis 72  
 394 ACTTGTGCTTAACCTTCTATGGCCATACCTAATAGCATGTTGATCACA 345  
 72 exGluGlyGlyGluValGlyArgPheValGlyGlnValMetGluAla 88  
 344 GTGAAGGGGAGAGGAGGTGGAGATTCTGTAGGCGAGGTGATGGAAGCC 295  
 89 LysArgHisSerLysGluGlnGlnMetArgProTyThrArgPheArgTh 105  
 294 AAGAAAGTCTAAGGAGCAACAGATGAGGCTTACACGCTTTCGGAAC 245  
 105 rProGluProAspAsnHisTyAspPheCysLeuIlePro 118  
 244 CCTGAACCTGACATCATTACGACTTTTGCCTCATACCT 205

seq\_name: gb\_est2:W35893

seq\_documentation\_block:  
 LOCUS W35893 472 bp mRNA EST 14-MAY-1996  
 DEFINITION mc53909.r1 Soares mouse embryo NME13.5 14.5 Mus musculus cDNA  
 clone IMAGE:352288 5' similar to SW:HG74\_HUMAN Q00994 OVARIAN  
 GRANULOSA CELL 13.0 KD PROTEIN HGR74. [1] ;, mRNA sequence.

ACCESSION W35893  
 VERSION W35893.1 GI:1317733  
 KEYWORDS EST.  
 SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 472)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The WashU-HMIT Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:224088

Seq primer: ETPPrimer  
 High quality sequence stop: 441.

#### FEATURES

source  
 1..472  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="IMAGE:352288"  
 /clone\_lib="Soares mouse embryo NbME13.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACCAATCTGAAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTT  
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
 State Univ., from 2 ]; double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 pT7T3 vector. Library went through one round of  
 normalization, and was constructed by Bento Soares and  
 M.Fatima Bonaldo."  
 BASE COUNT 149 a 95 c 157 g 71 t  
 ORIGIN

alignment\_scores:  
 Quality: 425.00 Length: 97  
 Ratio: 4.722 Gaps: 0  
 Percent Similarity: 92.784 Percent Identity: 85.567

alignment\_block:  
 US-09-327-750b-34 x W35893

Align seg 1/1 to: W35893 from: 1 to: 472

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17  
 |||||  
 181 ATGGCATCCAAATTTAAACAAGTCATCTGGATCTCCTGCTGGAGAAAGA 230  
 |||||  
 17 pLysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34  
 |||||  
 231 CAAAAAAGACAAAAGGTGGGAAGCCCTCCAAACAAGAGTGGAAGAAGAC 280  
 |||||  
 34 erHishisLeuGluGluValGluAsnLysLysProGlyGlyAsnValArg 50  
 |||||  
 281 CCCACCATCTGGAAGAGGTGGAACAAGAGCTGGGGGAATGTCCGA 330  
 |||||  
 51 ArgLysValArgArgLeuValProAsnPheLeuTrpAlaIleProAsnAr 67  
 |||||  
 331 AGGAAAGTCAGGCGACTTGTGCTTCTCTGGGCCATACCAATAG 380  
 |||||  
 67 gHisValAspHisSerGluGlyGlyGluGluValGlyArgPheValGlyG 84  
 |||||  
 381 GCATGTTGATCGCAATGAGGGGACAGGATGTTGGAGATTGTAGTGC 430  
 |||||  
 84 'lnValMetGluAlaLysArgHisSerLysGluGlnGlnMet 97  
 |||||  
 431 AGGGAACAGAAAGTCAAGAGAAAGACTACGAGCAGCAGGCTG 471

seq\_name: gb\_est2.W54487

seq\_documentation\_block:  
 LOCUS W54487 465 bp mRNA EST 03-JUN-1996

#### DEFINITION

md09c11.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA  
 clone IMAGE:367892 5', mRNA sequence.

ACCESSION W54487  
 VERSION W54487.1 GI:1355547  
 KEYWORDS EST.

#### SOURCE

house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 465)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilton,R. and  
 Waterston,R.

#### TITLE

The WashU-HHMI Mouse EST Project

#### JOURNAL

Unpublished (1996)

#### COMMENT

Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:229324

Seq primer: ETPPrimer

High quality sequence stop: 328.

#### FEATURES

source

1..465  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="IMAGE:367892"  
 /clone\_lib="Soares mouse embryo NbME13.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACCAATCTGAAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTT  
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
 State Univ., from 2 ]; double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 pT7T3 vector. Library went through one round of  
 normalization, and was constructed by Bento Soares and  
 M.Fatima Bonaldo."  
 BASE COUNT 147 a 90 c 156 g 72 t  
 ORIGIN

#### alignment\_scores:

Quality: 410.00 Length: 99  
 Ratio: 4.505 Gaps: 0  
 Percent Similarity: 91.919 Percent Identity: 84.848

#### alignment\_block:

US-09-327-750b-34 x W54487

Align seg 1/1 to: W54487 from: 1 to: 465

1 MetAlaSerLysValLysGlnValIleLeuAspLeuThrValGluLysAs 17  
 |||||  
 168 ATGGCATCCAAATTTAAACAAGTCATCTGGATCTCCTGCTGGAGAAAGA 217  
 |||||  
 17 pLysLysAsnLysLysGlyGlyLysAlaSerLysGlnSerGluGluGlu 34  
 |||||  
 218 CAAAAAAGACAAAAGGTGGGAAGCCCTCCAAACAAGAGTGGAAGAAGAC 267



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2002, 15:42:51 ; Search time 144 Seconds  
(without alignments)  
4167.553 Million cell updates/sec

Title: US-09-327-750D-28

Perfect score: 700

Sequence: 1 acagcgtctgcccagcagc.....ggggcttggtgcccagtga 700

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428652619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

- 1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	693.6	99.1	700	22	AAF23528	Mouse NADE DNA. M
2	307.4	43.9	891	22	AAF23529	Human NADE DNA. H
3	284.6	40.7	917	21	AAF21748	Human breast and o
4	253.6	36.2	532	21	AAC01005	Human secreted pro
5	192.2	27.5	485	21	AAC01004	Human secreted pro
6	147.2	21.0	187	21	AA87147	Rat hepatocyte car
7	127.2	18.2	426	22	AA82586	Human Ovarian tumo
8	80.2	11.5	865	22	AAH03517	Human cDNA clone (
9	80.2	11.5	1229	22	AAH13750	Human cDNA sequenc
10	80.2	11.5	1364	22	AAC85548	cDNA encoding CD1F
11	58	8.3	401	21	AAC01003	Human secreted pro

12	56.4	8.1	792	22	AAH45143	Human brain expres
13	55.8	8.0	767	22	AAH75810	Human x chromosome
14	55.8	8.0	898	22	AAF59611	Human cell cycle a
15	55.6	7.9	662	21	AAC03880	Human secreted pro
16	55.6	7.9	692	21	AAC10889	Human secreted pro
17	55.6	7.9	698	21	AAC03879	Human secreted pro
18	55.6	7.9	858	22	AAI60367	Human polynucleoti
19	54.2	7.7	862	22	AAI58581	Human polynucleoti
20	50.4	7.2	298	21	AAC06182	Human secreted pro
21	44.6	6.4	5552	21	AAAS8309	pIP/PyCSP.1 plasm
22	42.2	6.0	128	21	AAC12100	Human secreted pro
c 23	41.4	5.9	830	22	AAI22045	Probe #11978 for g
c 24	41.4	5.9	830	22	AAI47337	Probe #16023 used
c 25	41.4	5.9	830	22	AAI07740	Probe #7731 used t
26	41	5.9	279	21	AAC06181	Human secreted pro
27	40.8	5.8	381	22	AAI12579	Probe #2512 for ge
28	40.8	5.8	381	22	AAI33926	Probe #2612 used t
29	40.8	5.8	381	22	AAI02485	Probe #2476 used t
30	40.4	5.8	10732	21	AAAI0594	Gene encoding a su
31	39.2	5.6	1635	22	AAI22780	Probe #12713 for g
32	39.2	5.6	1635	22	AAI48082	Probe #16768 used
33	39.2	5.6	1635	22	AAI08454	Probe #8445 used t
34	39.2	5.6	1973	22	AAI13556	Probe #3489 for ge
35	39.2	5.6	1973	22	AAI34918	Probe #3604 used t
36	39.2	5.6	1973	22	AAI03446	Human secreted pro
37	38.8	5.5	1452	21	AAA26304	Human cDNA sequenc
38	38.6	5.5	2551	22	AAH17742	Arabidopsis thalia
39	38.4	5.5	1165	21	AAC38965	Human ORF ORF1926
c 40	38.4	5.5	1183	21	AAC76371	Human ORF ORF1926
c 41	38.4	5.5	1733	22	AAAS0359	Human diagnostic a
42	37.6	5.4	439	22	AAI22718	Probe #12651 for g
43	37.6	5.4	439	22	AAI48016	Probe #16702 used
44	37.6	5.4	439	22	AAI08383	Probe #8374 used t
45	37.6	5.4	1709	21	AAFI5695	Human prostate can

ALIGNMENTS

RESULT 1  
AAF23528  
ID AAF23528 standard; DNA; 700 BP.  
AC AAF23528;  
XX  
DT 22-MAR-2001 (first entry)  
XX  
DE Mouse NADE DNA.  
XX  
KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;  
KW neurogenetic disease; NF-kappaB; ds.  
XX  
OS Mus sp.  
XX  
PN WO200075278-A2.  
XX  
PD 14-DEC-2000.  
XX  
PF 07-JUN-2000; 2000WO-US15621.  
XX  
PR 07-JUN-1999; 99US-0327750.  
XX  
PA (UYCO-) UNIV COLUMBIA NEW YORK.  
XX  
PI Sato T.  
XX  
DR WPI; 2001-061707/07.  
XX  
PT New p75-neurotrophin receptor-associated cell death executor (NADE) and  
PT the gene encoding NADE, useful for modulating the activity of p53NTR  
XX and for detecting neurodegenerative diseases -  
PS Claim 12; Fig 1; 134pp; English.











CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 1229 BP; 351 A; 243 C; 296 G; 339 T; 0 other;

Query Match 11.5%; Score 80.2; DB 22; Length 1229;  
Best Local Similarity 56.2%; Pred. No. 1.8e-12;  
Matches 269; Conservative 0; Mismatches 153; Indels 57; Gaps 4;

QY 143 gccagaatagcaaccaggaaacaaatctcatcgtgccaatgtccaccaggaaacga 202  
DB 196 gtccaaagaggaactagcgccaacaatctcaacggggaataatgcccaacagaaacga 255  
QY 203 agagctggagcag---ccctgcagaatggacagaaacacccctgtggaggaggtga 259  
DB 256 agagggggagcagggcccccacgcagaatgaagaagaatcccgccatttgggggggtga 315  
QY 260 gggccaccagcctgtcgaaacaaacaaacaaacacacacacacacacacacacac 319  
DB 316 aggcagaagcctggaggaaatatac----- 341  
QY 320 ccgaagagcggagcctgcgcacccctcacttccgatggccattcccaacaggca 379  
DB 342 --ggcggggcgagtaggcgactgtccctaatttcgatggccatacctaataaggca 399  
QY 380 gatgaatcacgggttgggtggagatggagatgatgaaatgttcagtgagagatgag 439  
DB 400 tattgagcaca-----tgaagcagagatgatgagaaggttggtaggcagatgat 453  
QY 440 agagatccggagaagcttaggagctacagctgagaatgtctacgcactctatggg 499  
DB 454 ggaaatcaagagaagactagggaacagcagatgaggcactatagcgttccaaa----- 509  
QY 500 ggaagctgtctaccaccacgacatcacatgatgaattctgccttgccttgacttcggtc 559  
DB 510 -----ctcctgaacctgacaacattatgactttgctcactacactgaactctaa 561  
QY 560 attccccctgagatcactatgtactcccgctgtagcccttccctcgcatttct 618  
DB 562 agtttctgctgaggttaattgtgaacactgtcttacaagctgtattttgtgatttact 620

## RESULT 10

AAC85548  
ID AAC85548 standard; cDNA; 1364 BP.

XX AAC85548;

XX AAC85548;

XX 04-JUN-2001 (first entry)

DE cDNA encoding CDIFF-4, Incyte ID No. 1990956CB1.

XX Human; cell differentiation; CDIFF; agonist; antagonist; epilepsy;

KW cell proliferation; Alzheimer's disease; schizophrenia disorder;

KW arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 250..612

XX /tag= a

XX /product= "CDIFF-4"

XX WO200119860-A2.

XX 22-MAR-2001.  
XX  
XX 14-SEP-2000; 2000WO-US25435.  
XX  
XX 15-SEP-1999; 99US-0154140.  
XX 06-DEC-1999; 99US-0169155.  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Tang YT, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;  
XX Baughn MR, Lu DAM, Bandman O, Shih LL, Patterson C;  
XX WPI; 2001-211447/21.  
XX P-PSDB; AAB47126.  
XX Isolated polypeptides and polynucleotides involved in cell  
XX differentiation are used for treatment, prevention and diagnosis of  
XX cell proliferative, developmental and neurological disorders e.g.  
XX cancer and Alzheimer's disease -  
XX  
XX Claim 5; Page 121; 137pp; English.  
XX The sequences given in AAC85545-72 encode human polypeptides involved  
XX in cell differentiation (CDIFF). CDIFF polypeptides and agonists of  
XX these are used to treat a disease or condition associated with  
XX decreased expression of functional CDIFF. An antagonist of CDIFF is  
XX used to treat a disease or condition associated with over expression  
XX of functional CDIFF. CDIFF polypeptides may be used for the treatment,  
XX prevention and diagnosis of cell proliferative, developmental and  
XX neurological disorders, such as Alzheimer's disease, schizophrenia  
XX disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus  
XX and epilepsy. The CDIFF-4 sequence is homologous to Mus musculus  
XX REX-3. This sequence maps to chromosome 1 within the interval from  
XX 152.2 to 157.4 centiMorgans, to chromosome 3 within the interval from  
XX 157.4 to 158.0 centiMorgans, and to the X chromosome within the interval  
XX from 104.9 to 150.3 centiMorgans.  
XX  
XX Sequence 1364 BP; 411 A; 269 C; 322 G; 362 T; 0 other;

Query Match 11.5%; Score 80.2; DB 22; Length 1364;  
Best Local Similarity 56.2%; Pred. No. 1.8e-12;  
Matches 269; Conservative 0; Mismatches 153; Indels 57; Gaps 4;

QY 143 gccagaatagcaaccaggaaacaaatctcatcgtgccaatgtccaccaggaaacga 202

DB 255 gtccaaagaggaactagcgccaacaatctcaacggggaataatgcccaacagaaacga 314

QY 203 agagctggagcag---ccctgcagaatggacagaaacacccctgtggaggaggtga 259

DB 315 aggaaggaggcagggcccccacgcagaatgaagaagaatcccgccatttgggggggtga 374

QY 260 gggccaccagcctgtcgaaacaaacaaacaaacacacacacacacacacacacac 319

DB 375 aggcagaagcctggaggaaatatac----- 400

QY 320 ccgaagagcggcctgcgcacccctcacttccgatggccattcccaacaggca 379

DB 401 --ggcggggcgagtaggcgactgtccctaatttcgatggccatacctaataaggca 458

QY 380 gatgaatcacgggttgggtggagatggagatgatgaaatgttcagtgagagatgag 439

DB 459 tattgagcaca-----tgaagcagagatgatgagaaggttggtaggcagatgat 512

QY 440 agagatccggagaagccttaggagctacagctgagaatgtctacgcactctatggg 499

DB 513 ggaaatcaagagaagactagggaacagcagatgaggcactatgcgttccaaa----- 568

QY 500 ggaagctgttaaccaccacgacatcacatgatgaattctccttatgcttgccttgctc 559

DB 569 -----ctcctgaacctgacaacccattatgactttgctcactacactgaatcctaa 620





AC	AAH75810;
XX	
DT	17-OCT-2001 (first entry)
DE	
XX	Human X chromosome linked gene expression protein 14 coding sequence.
XX	
XX	Human; X chromosome linked gene expression protein 14; cancer;
KW	HIV infection; cytostatic; anti-HIV; chromosome X; ss.
XX	
OS	Homo sapiens.
XX	
PN	CN1296969-A.
XX	
PD	30-MAY-2001.
XX	
PF	23-NOV-1999; 99CN-0124078.
XX	
PR	23-NOV-1999; 99CN-0124078.
XX	
PA	(SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
PI	Mao Y, Xie Y;
XX	
DR	WPI: 2001-483897/53.
DR	P-PSDB; AAG66407.
XX	
PT	Polypeptide-human X chromosome linked gene expression protein 14 and
PT	polynucleotide for coding said polypeptide
XX	
PS	Claim 6; Page 24 (Disclosure); 31pp; Chinese.
XX	
CC	The present sequence is the coding sequence for human X chromosome linked
CC	gene expression protein 14. The protein and coding sequence are useful
CC	for treating diseases e.g. cancer and HIV infection.
XX	
SQ	Sequence 767 BP; 224 A; 143 C; 215 G; 185 T; 0 other;
Query Match 8.0%; Score 55.8; DB 22; Length 767;	
Best Local Similarity 53.7%; Pred. No. 8.4e-06;	
Matches 139; Conservative 0; Mismatches 117; Indels 3; Gaps 1	
QY	316 accacgaagaggcaggctgcgcgacatgccctaacttccgatggcgcatcccccaaca 375 
Db	255 accgtagcggttcgcgcttaggcagcccacctgcagtatatggacataatgcata 314 
QY	376 ggcagatgaatgacgggttggtggagatgagatgatggaattctcatggaggaga 435 
Db	315 ggcttgagagccacagcaaggatgagagaggagaatatggaagattggggaggagg 374 
QY	436 taagacagatccgagaaagcttagggactcacagctgaaaaattgtctacgcattcta 495 
Db	375 tgagacagctgatgaaaagctggagggaagacagtgagtcactgctggggcagttca 434 
QY	496 tgggggagctgtctaacaccacgacatcaccatgatgaattctgcttatgccttgacttc 555 
Db	435 gcactgatccccct---caccatgaccatcagatgatttbgccttatccctgaatcc 491 
QY	556 ggtcattccccctcctgagat 574
Db	492 tgatgggtttccctggaqgt 510

Cell cycle and proliferation protein; CCYPR; human; agonist;  
antagonist; gene therapy; detection; gene therapy;  
transgenic animal disease model; immune disorder;  
developmental disorder; cell signalling disorder;  
cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
arteriosclerosis; asthma; allergy; diabetes mellitus;  
menstrual cycle disorder; bacterial infection; ss.

Homo sapiens.

WO200107471-A2.

01-FEB-2001.

21-JUL-2000; 2000WO-US19948.

21-JUL-1999; 99US-0145075.  
08-SEP-1999; 99US-0153129.  
10-NOV-1999; 99US-0164647.

(INCY-) INCYTE GENOMICS INC.

Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;  
Aizimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;  
WPI; 2001-112727/12.

P-PSDB; AAB60474.

Human cell cycle and proliferation proteins and polynucleotides are  
used to treat, diagnose and prevent immune, developmental and cell  
signaling disorders and cell proliferative disorders including cancer -

Claim 5; Page 181-182; 205pp; English.

Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human  
cell cycle and proliferation proteins (CCYPR). AAB60453-AAB60506.  
CCYPR and agonists of CCYPR are used to treat diseases or conditions  
associated with decreased expression of functional CCYPR, while CCYPR  
antagonists are used to treat diseases or conditions associated with  
overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
that specifically bind to CCYPR, and in drug screening methods to  
identify compounds that modulate the activity of CCYPR. CCYPR  
nucleotides can be used to generate transgenic animal models of human  
disease, and can be used in gene therapy in target cells with genetic  
abnormalities with respect to the expression of CCYPR for the  
treatment or prevention of a disorder associated with CCYPR.  
Diseases which can be diagnosed, treated and prevented using CCYPR  
proteins, nucleic acids, agonists or antagonists include immune,  
developmental and cell signalling disorders, and cell proliferative  
disorders including cancer. Specific examples of these disorders  
include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
diabetes mellitus, disorders of the menstrual cycle and infections  
caused by bacteria.

Sequence 898 BP: 250 A; 186 C; 251 G; 211 T; 0 other;

	Query Match	8.0%	Score 55.8;	DB 22;	Length 898;
	Best Local Similarity	53.7%;	Pred. No. 9e-06;		
	Matches 139;	Conservative 0;	Mismatches 117;	Indels 3;	Gaps 1;
Qy	316	accaccgaagaggcgagctgcgcgacctgcgcctaaacttcgcattcgatggcgcaattccccaaca	375		
Db	408	accgtaggcggttcgcgcttagggcagccactcctgcagatagatggagacataatgcata	467		
Qy	376	ggcagatgaatgcagggttggtggagatggagatgatgataatgaaatgttcattcaggagaga	435		
Db	468	ggcttgagagccacagcgagatgagaggagaaatgaaaggaattgggggagagg	527		
Qy	436	tgaagaagatccgagaagaagcttaaggagctacagctgaagaattgtctacqcatccta	495		

Db	528	tggagacagctga	tggaaaagctgagggaaaagcagttgagtcatagtctgcgggcagtc	ca	587
QY	496	tgggggagctgtctaa	ccaccacagatcaccatgatgaattctgccttatgccttgacttc		555
Db	588	gcactgatccctc	---caccatgaccatcacagatgagtttgccttatgcctgaatcc		644
QY	556	ggtcattccccctgagat		574	
Db	645	tgatggtttccctgaagt		663	

RESULT 15

AAC03880

ID AAC03880 standard; cDNA; 662 BP.

AAC03880;

DT 06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 3878.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW  
gene therapy; chromosome mapping; ss.  
KW

OS Homo sapiens.

PN EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

PA (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J:

DR WPI; 2000-500381/45.

DR P-PSDB; AAG03874.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 3878; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. An ORF has been identified within the sequence. The 5' ESTs were prepared from total human RNAs or poly(A)<sup>+</sup> RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

SQ Sequence 662 BP; 177 A: 144 C: 192 G: 137 T: 12 other:

Query Match 7.9%: Score 55.6: DB 21: Length 662:

Query Match: 7.5%, score 33.0, DB  
Best Local Similarity 61.28; pred. No. 9e-06;

Matches	104;	Conservative	1;	Mismatches	62;	Indels	3;	Gaps	1;
---------	------	--------------	----	------------	-----	--------	----	------	----

Qy 405 ggagatgatatggaatgttcatggaggagatgagagagatccggagaaagcttagggag 464

[illegible]

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2002, 16:36:39 ; Search time 144 Seconds  
(without alignments)  
5304.700 Million cell updates/sec

Title: US-09-327-750D-29

Perfect score: 891

Sequence: 1 accccatccccctctctat.....aataagcaatttaaaagc 891

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	889.4	99.8	891	22	AAF23529 Human NADE DNA. H
2	564.4	63.3	917	21	AAF21748 Human breast and o
3	343	38.5	532	21	AAC01005 Human secreted pro
4	310.6	34.9	700	22	AAF23528 Mouse NADE DNA. M
5	298.8	33.5	485	21	AAC01004 Human secreted pro
6	273.8	30.7	426	22	AAH82586 Human ovarian tumo
7	160.8	18.0	169	16	AAAT21034 Human gene signatu
8	150.8	16.9	187	21	AAH87147 Rat hepatocyte car
9	141.2	15.8	865	22	AAH03517 Human cDNA clone (
10	141.2	15.8	1229	22	AAH13750 Human cDNA sequenc
11	141.2	15.8	1364	22	AAC85548 cDNA encoding CDIF

12	108.2	12.1	792	22	AAH45143 Human brain expres
13	104	11.7	898	22	AAF59611 Human cell cycle a
14	102.4	11.5	767	22	AAF75810 Human x chromosome
15	102.4	11.5	862	22	AAI58581 Human polynucleoti
16	95.4	10.7	858	22	AAI50367 Human polynucleoti
17	91.4	10.3	692	21	AAC10889 Human secreted pro
18	89.8	10.1	662	21	AAC03880 Human secreted pro
19	89.8	10.1	698	21	AAC03879 Human secreted pro
c 20	81	9.1	936	22	AAF58254 Oligonucleotide D1
c 21	81	9.1	936	22	AAF58254 Oligonucleotide D1
c 22	81	9.1	936	22	AAF58257 Oligonucleotide D1
c 23	81	9.1	936	22	AAF58259 Oligonucleotide D2
c 24	81	9.1	936	22	AAF58262 Oligonucleotide D1
c	81	9.1	938	22	AAF58255 Oligonucleotide D1
25	74.6	8.4	1197	21	AAH72409 Human nucleic acid
26	68.2	7.7	401	21	AAC01003 Human secreted pro
27	67.4	7.6	936	22	AAF58252 Oligonucleotide D1
28	67.4	7.6	936	22	AAF58254 Oligonucleotide D1
29	67.4	7.6	936	22	AAF58257 Oligonucleotide D1
30	67.4	7.6	936	22	AAF58259 Oligonucleotide D2
31	67.4	7.6	936	22	AAF58252 Oligonucleotide D1
32	67.4	7.6	936	22	AAF58254 Oligonucleotide D1
33	67.4	7.6	938	22	AAF58255 Oligonucleotide D1
34	66.4	7.5	128	21	AAC12100 Human secreted pro
c 35	64.8	7.3	397	22	AAI14082 Probe #4015 for ge
c 36	64.8	7.3	397	22	AAI35463 Probe #4149 used t
c 37	64.8	7.3	397	22	AAI03935 Probe #3926 used t
38	58.6	6.6	1170	22	AAH68557 Human protein HP10
39	58.6	6.6	1250	22	AAI57960 Human polynucleoti
40	58.6	6.6	1551	22	AAI59746 Human polynucleoti
41	56.4	6.3	396	21	AAH30950 Human colon cancer
42	55.4	6.2	1309	21	AAC98961 Human pancreatic c
c 43	52.8	5.9	285	22	AAI21436 Probe #11369 for g
c 44	52.8	5.9	285	22	AAI46725 Probe #15411 used
c 45	52.8	5.9	285	22	AAI07131 Probe #7122 used t

ALIGNMENTS

RESULT	1
AAF23529	AAF23529 standard; DNA; 891 BP.
ID	AAF23529 standard; DNA; 891 BP.
XX	
AC	AAF23529;
XX	
DT	22-MAR-2001 (first entry)
XX	
DE	Human NADE DNA.
XX	
KW	Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;
KW	neurogenetic disease; NF-kappaB; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200075278-A2.
XX	
PD	14-DEC-2000.
XX	
PF	07-JUN-2000; 2000WO-US15621.
XX	
PR	07-JUN-1999; 99US-0327750.
XX	
PA	(UYCO ) UNIV COLUMBIA NEW YORK.
XX	
PI	Sato T.
XX	
DR	WPI; 2001-061707/07.
XX	
PT	New p75-neurotrophin receptor-associated cell death executor (NADE) and
PT	the gene encoding NADE, useful for modulating the activity of p75NTR
XX	and for detecting neurodegenerative diseases -
PS	Disclosure; Fig 1; 134pp; English.

XX The present invention relates to a purified polypeptide capable of  
CC binding neurotrophin receptor (p75-NTR). The invention is useful for  
CC binding and modulating the activity of p75NTR. The peptide mediates  
CC NGF-induced apoptosis, which plays an important role in neurogenetic  
CC diseases. The peptide of the invention and p75NTR are useful for  
CC inhibiting NF-kappaB activation in a cell or a subject, for inducing  
CC caspase-2 and caspase-3 activity to cleave poly (ADP-ribose) polymerase  
CC and fragment nuclear DNA in a cell by co-expression of (1) and p75-NTR.  
XX  
SQ Sequence 891 BP; 251 A; 182 C; 224 G; 234 T; 0 other;

Query Match 99.8%; Score 889.4; DB 22; Length 891;  
Best Local Similarity 99.9%; Pred. No. 5.9e-287;  
Matches 890; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 accccatccccactctctatcacggtctctccatttggctgcctcaagctctgggaaag 60  
Db 1 accccatccccactctctatcacggtctctccatttggctgcctcaagctctgggaaag 60  
Qy 61 aatcccggaacgaaataatggtgggttggggaagggagtaaggaggaagaaagctgga 120  
Db 61 aatcccggaacgaaataatggtgggttggggaagggagtaaggaggaagaaagctgga 120  
Qy 121 gggaggggtcttaattgagggcccgtagagacgcgcgaactcttaagctgggaaaaa 180  
Db 121 gggaggggtcttaattgagggcccgtagagacgcgcgaactcttaagctgggaaaaa 180  
Qy 181 acgaataataaaatcccttgatatacagggtctgaactctgctggtcagagcacaagc 240  
Db 181 acgaataataaaatcccttgatatacagggtctgaactctgctggtcagagcacaagc 240  
Qy 241 attcagtcctctccttgccttcttacttcttacttcttacttcttacttcttactt 300  
Db 241 attcagtcctctccttgccttcttacttcttacttcttacttcttacttcttactt 300  
Qy 301 aaaaatctcatatgcaaatatccaccaggaagaaacgaagagatggagcagctatgcaga 360  
Db 301 aaaaatctcatatgcaaatatccaccaggaagaaacgaagagatggagcagctatgcaga 360  
Qy 361 atggagaggaagacgcctcttggagaggtggaagccaccagcctcaggaataatgcac 420  
Db 361 atggagaggaagacgcctcttggagaggtggaagccaccagcctcaggaataatgcac 420  
Qy 421 ggggacaggtcgcagacttcccctaatatttcagatggggccataccccaataggcagatca 480  
Db 421 ggggacaggtcgcagacttcccctaatatttcagatggggccataccccaataggcagatca 480  
Qy 481 atgatggatgggtggagatgagatgatgataatggaataattcatggagagatgagagaaa 540  
Db 481 atgatggatgggtggagatgagatgatgataatggaataattcatggagagatgagagaaa 540  
Qy 541 tcagaagaaacttaggagctgagtgaggaattgctgcgtatccttatgggggagc 600  
Db 541 tcagaagaaacttaggagctgagtgaggaattgctgcgtatccttatgggggagc 600  
Qy 601 tcttaataacacatgacatcatgatgaatttgccttatgccttgactgactgactgactt 660  
Db 601 tcttaataacacatgacatcatgatgaatttgccttatgccttgactgactgactgactt 660  
Qy 661 tcatgagataataactgatacccgctgttttctttcttcttcttcttcttcttcttcttct 720  
Db 661 tcatgagataataactgatacccgctgttttctttcttcttcttcttcttcttcttcttct 720  
Qy 721 ctttactgacggttggctggaacccctatgtttttccatgtgtcagtggtggtctgtg 780  
Db 721 ctttactgacggttggctggaacccctatgtttttccatgtgtcagtggtggtctgtg 780  
Qy 781 ttgcaagcttctatttgaagattgaccttgcactcagtgtaagtttctgtcagcagtagt 840  
Db 781 ttgcaagcttctatttgaagattgaccttgcactcagtgtaagtttctgtcagcagtagt 840

Qy 841 ttaccatttcattgcatggaaaaattttaaaagccaataaagcaatttaaaagc 891  
Db 841 ttaccatttcattgcatggaaaaattttaaaagccaataaagcaatttaaaagc 891

## RESULT 2

AAF21748

ID AAF21748 standard; DNA; 917 BP.

XX AAF21748;

XX 27-MAR-2001 (first entry)

XX Human breast and ovarian cancer associated antigen gene SEQ ID 135.

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antineoplastic; antitumor; antitumor; antitumor;  
KW antidiabetic; antineoplastic; antitumor; antitumor; antitumor;  
KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
KW cardiovascular disorder; wound healing; neurological disease; ds.

OS Homo sapiens.

PN WO200055173-A1.

PD 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05881.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

XX WPI; 2000-611515/58.

DR P-PSDB; AAB58845.

XX New human breast and ovarian cancer associated gene sequences and the  
PT polypeptides encoded by these genes, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -

XX Claim 1; Page 581-582; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
CC associated with breast and ovarian cancer. Included in the invention are  
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
CC isolation and characterisation of the DNA and protein sequences of the  
CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
CC antidiabetic; antineoplastic; antitumor; antitumor; antitumor;  
CC antibacterial; antifungal; antiparasitic and cardiant activity. The  
CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
CC particularly breast and ovarian cancer. The nucleic acid sequences,  
CC proteins, agonists and antagonists may also be used in the diagnosis,  
CC prevention and treatment of immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemia; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases.

XX Sequence 917 BP; 228 A; 203 C; 270 G; 211 T; 5 other;

Query Match 63.3%; Score 564.4; DB 21; Length 917;  
Best Local Similarity 99.8%; Pred. No. 2.5e-159;

Matches 565; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 326 ccaggaaacaaagagatggagcagcctatgcagaatggagagagagcgcctttggg 385  
|||||  
Db 305 ccaggaaacaaagagatggagcagcctatgcagaatggagagagagcgcctttggg 364  
Qy 386 agggaggtgaagccaccagcctgcaggaaatgcagcgggagcgcctccgacttgcgcc 445  
|||||  
Db 365 agggaggtgaagccaccagcctgcaggaaatgcagcgggagcgcctccgacttgcgcc 424  
Qy 446 taatttgcagtgggccatacccaatagcagatcaatgatgggatgggtgagatggaga 505  
|||||  
Db 425 taatttgcagtgggccatacccaatagcagatcaatgatgggatgggtgagatggaga 484  
Qy 506 tgatatgaatattcatctggaggagatgagagaaatcagaagaaacttagggagctgca 565  
|||||  
Db 485 tgatatgaatattcatctggaggagatgagagaaatcagaagaaacttagggagctgca 544  
Qy 566 gtgagggaattgtctgccttcttcttggggagcctctctaatcaccatgaccatcatga 625  
|||||  
Db 545 gtgagggaattgtctgccttcttcttggggagcctctctaatcaccatgaccatcatga 604  
Qy 626 tgaatttgccttgccttgcactcctccatttgcattatgcagattaatcattgattccc 685  
|||||  
Db 605 tgaatttgccttgccttgcactcctccatttgcattatgcagattaatcattgattccc 664  
Qy 686 gctgtttctttcttctgcatttccataatgccttactgacgcttcttgccttgcctgaac 745  
|||||  
Db 665 gctgtttctttcttgcatttccataatgccttactgacgcttcttgccttgcctgaac 724  
Qy 746 cctatgttatttccatgtgtcaagtgggtcttcttgccttgccttcttgccttgcctgaac 805  
|||||  
Db 725 cctatgttatttccatgtgtcaagtgggtcttcttgccttgccttcttgccttgcctgaac 784  
Qy 806 ctttgcactcagtgtaagtcttgcagcagtagtttccaccatttgcagtgagaaattt 865  
|||||  
Db 785 ctttgcactcagtgtaagtcttgcagcagtagtttccaccatttgcagtgagaaattt 844  
Qy 866 aaagccaaataagcaatttaaaagc 891  
|||||  
Db 845 aaagccaaataagcaatttaaaagc 870

## RESULT 3

AAAC01005  
ID AAC01005 standard; cDNA; 532 BP.  
XX  
AC AAC01005;

XX  
DT 06-OCT-2000 (first entry)

XX  
DE Human secreted protein 5' EST, SEQ ID NO: 1003.

XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

XX  
OS Homo sapiens.

XX  
PN EP1033401-A2.

XX  
PD 06-SEP-2000.

XX  
PF 21-FEB-2000; 2000EP-0200610.

XX  
PR 26-FEB-1999; 99US-0122487.

XX  
PA (GEST ) GENSET.

XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX  
XX WPI; 2000-500381/45.

XX  
DR P-PSDB; AAG00999.

XX  
XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
PS Claim 1; SEQ ID 1003; 71pp + CD-ROM; English.  
XX

CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.

XX Sequence 532 BP; 151 A; 121 C; 163 G; 91 T; 6 other;

Query Match 38.5%; Score 343; DB 21; Length 532;  
Best Local Similarity 98.6%; Pred. No. 6e-93;  
Matches 340; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 281 agaaaaaacacagaaaaaaatctctatcgtgcaaatatttcaccaggaacgaaga 340  
|||||

Db 188 agaaaaaacacagaaaaaaatctctatcgtgcaaatatttcaccaggaacgaaga 247  
|||||

Qy 341 atgtagagcagcctatgcagaatggagagagacccgcttggagagggtgaagcca 400  
|||||

Db 248 gatgagcagcctatgcagaatggagagagacccgcttggagagggtgaagcca 307  
|||||

Qy 401 ccagcctgcagaaatcgacggggacagctcgcgacttgccttaatttcgatgggc 460  
|||||

Db 308 ccagcctgcagaaatcgacggggacagctcgcgacttgccttaatttcgatgggc 367  
|||||

Qy 461 catacccaatagcagagatcaatgatgggtgggtgagatgagatgataatgaaatatt 520  
|||||

Db 368 catacccaatagcagagatcaatgatgggtgggtgagatgagatgataatgaaatatt 427  
|||||

Qy 521 catgagagatgagagaaatcagaagaaacttagggagctgcagttgagaaattgtct 580  
|||||

Db 428 catggaggagatgagagaaacttagggagctgcagttgagaaattgtct 487  
|||||

Qy 581 gcgtatccttgggggagctctctaatcaccatgaccatcatga 625  
|||||

Db 488 gcgtatccttgggggagctctctaatcaccatgaccatcatga 532  
|||||

## RESULT 4

AAAF23528

ID AAFA23528 standard; DNA; 700 BP.

XX  
AC AAFA23528;

XX  
DT 22-MAR-2001 (first entry)

XX  
DE Mouse NADE DNA.

XX  
KW Neurotrophin receptor; p75-NTR; NGF-induced apoptosis;

XX  
KW neurogenetic disease; NF-kappab; ds.

XX  
OS Mus sp.

XX  
PN WO200075278-A2.

XX  
XX 14-DEC-2000.

XX  
PF 07-JUN-2000; 2000WO-US15621.



QY 566 gttgaggaattgtgtcgctatcttattggtgggagctctctataatcaccatgaccatcatga 625  
:|||||  
Db 426 ktgaggaattgtgtcgctatcttattggtgggagctctctataatcaccatgaccatcatga 485

## RESULT 6

AAH82586  
ID AAH82586 standard; cDNA; 426 BP.

XX AC AAH82586;

XX DT 25-SEP-2001 (first entry)

XX DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:210.

XX KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;

XX KW immunogenic; vaccine; ss.

XX XX Homo sapiens.

XX OS WO200151513-A2.

XX PN 19-JUL-2001.

XX PD 16-JAN-2001; 2001WO-US01575.

XX PR 14-JAN-2000; 2000US-0176722.

XX PA (CORI-) CORIXA CORP.

XX PI Algate PA;

XX PI WPI; 2001-425866/45.

XX DR Novel ovarian tumor proteins, and nucleic acids encoding them, used to  
PT treat and diagnose cancers, particularly ovarian cancer -

XX PS Claim 5; Page 100; 338pp; English.

XX CC AAH82377 to AAH83878 represent human ovarian tumour-associated  
CC polynucleotide sequences which encode ovarian tumour proteins. The  
CC ovarian tumour protein and polynucleotide sequences have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. The  
CC ovarian tumour proteins and polynucleotides can be used to inhibit  
CC the development of cancer, particularly ovarian cancer. They can also  
CC be used to diagnose the onset and progression of cancer.

XX SQ Sequence 426 BP; 86 A; 106 C; 90 G; 137 T; 7 other;

Query Match 30.7%; Score 273.8; DB 22; Length 426;  
Best Local Similarity 97.5%; Pred. No. 3.2e-72;  
Matches 308; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

QY 552 cttagggagctgcagttgaggaattgtctgtatcttattggtgggagctctctaatcac 611  
|||||

Db 1 cttagggagctgcanttgaggaattgtctgtatcttattggtgggagctctctaatcac 60

QY 612 catgaccatcatgagaaatttgcttattgcttgccttgactctccattatcatgagatta 671  
|||||

Db 61 catgaccatcatgagaaatttgcttattgcttgccttgactctccattatcatgagatta 120

QY 672 atactgtgattcccgctgtttctttcttcttgcttcttccataatgcctttactgac 731  
|||||

Db 121 atactgtgattcccgctgtttctttcttcttgcttcttccataatgcctttactgac 179

QY 732 cgtttgtgtgaaccttatgttatttccatgtgtcaagtgggtcttctgtgt-tgccagctt 790  
|||

Db 180 cgtttgtgtgaaccttatgttatttccatgtgtcaagtgggtcttctgtgtgtcgcagctt 238

QY 791 ctatttgaagattgcttggcactcagtgtaagtgttctgtcagcagtagtttcaaccatt 850  
|||||

Db 239 ctatttgaagattgcttggcactcagtgtaagtgttctgtcagcagtagtttcaaccatt 298  
|||||

QY 851 tgcatggaaaaatttta 866  
|||||  
Db 299 tgcatggaaaaatttta 314

## RESULT 7

AAAT21034  
ID AAAT21034 standard; cDNA to mRNA; 169 BP.

XX AC AAAT21034;

XX DT 03-JUL-1996 (first entry)

XX DE Human gene signature HUMGS02316.

XX KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
human; cloning; mapping; non-biased library; diagnosis; detection;  
cell typing; abnormal cell function; ss.

XX OS Homo sapiens.

XX PN WO9514772-A1.

XX PD 01-JUN-1995.

XX PF 11-NOV-1994; 94WO-JP01916.

XX PR 12-NOV-1993; 93JP-0355504.

XX PA (MATS/) MATSUBARA K.  
(OKUB/) OKUBO K.

XX PI Matsubara K, Okubo K;

XX DR WPI; 1995-206931/27.

XX PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
for diagnosis of abnormal cell function, by preparing cDNA that  
reflects relative abundance of corresp. mRNA in specific human  
tissues

XX PS Claim 1; Page 786; 2245pp; Japanese.

XX CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in AAT19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

XX SQ Sequence 169 BP; 47 A; 30 C; 34 G; 58 T; 0 other;

Query Match 18.0%; Score 160.8; DB 16; Length 169;  
Best Local Similarity 98.8%; Pred. No. 1.7e-38;  
Matches 162; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 728 gatccgtttgtcgtgaacctatgtttatccatgtgtcaagtgggtcttctgttccag 787  
|||||

Db 1 gatccgtttgtcgtgaacctatgtttatccatgtgtcaagtgggtcttctgttccag 60

QY 788 ctctcatttgaagattgcctttgcactcagtgtaagtcttctgcagcagtagtttccacc 847  
|||||

Db 61 ctctattgaagattgcgttgcactcaatgaagtctctgcagcagtagttcaccc 120

Qy 848 atttgcattgaaataattaaagccaataagaacatttaaaagc 891  
|||||

Db 121 atttgcattgaaataattaaagccaataagaacatttaaaagc 164  
|||||

RESULT	8
AAA87147	
ID	AAA87147 standard; DNA; 187 BP.
XX	
AC	AAA87147;
XX	
DT	08-JAN-2001 (first entry)
XX	
DE	Rat hepatocyte carcinogenesis biomarker nucleic acid SEQ ID NO:71.
XX	
KW	Rat; phenobarbital; carcinogenesis marker; carcinogenesis; detection;
XX	identification; carcinogenic; probe; primer; ds.
XX	
OS	Rattus norvegicus.
XX	
PN	WO200044902-A2.
XX	
PD	03-AUG-2000.
XX	
PF	28-JAN-2000; 2000WO-US00503.
XX	
PR	29-JAN-1999; 99US-0118078.
XX	
PA	(SEAR ) SEARLE & CO G D.
XX	
PI	Bunch RT, Curtis SW, Rodi CP, Morris DL;
XX	
DR	WPI; 2000-505977/45.
XX	
PT	New nucleic acid encoding a carcinogenic biomarker, induced by
PT	phenobarbital treatment of rat hepatocytes, useful for identifying
PT	carcinogenic compounds -
XX	
PS	Claim 1; Page 73; 240pp; English.

RESULT 9

AAH03517  
ID AAH03517 standard; cDNA; 865 BP.

XX  
XX  
AC AAH03517;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:352.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
WI 2001-3118749/34.

DR WPI; 2001-3118749/34.

XX  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs .  
XX  
XX Claim 1; SEQ ID 352; 2537pp + CD ROM; English.  
PS  
XX The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 865 BP; 239 A; 177 C; 224 G; 221 T; 4 other:

182 ccttta 187







Db 423 ggattgggagggggtgagacagctgatgaaaaagctgagggaaagcagttgagtcata 482

Qy 577 gtcgcgtatccttatggggagctcttaaccacatgaccatcatgatgaatttggcc 636

Db 483 gtcgaggcagtcagcactgacccccc---tcaccatgaccatcatgatgatttggcc 539

Qy 637 ttatgcctgactcctgcattatcatgagatttaatactgtgattccgcgtgtttctt 696

Db 540 ttatgcctgactcctgatttccctaaagtattac--ggaaacagaccctgcttt 597

Qy 697 ttctctgactttcctaataatgcttttactgatccggtttgtgtgaaacctatattt 756

Db 598 cgaattacatgttcatgatgctgctgtgtgtgaaacctttacctgtcactgtttt--- 653

Qy 757 tccatgtgcaagtggtgtgtgtgctgagcttctatttgaagattgcttt 809

Db 654 -----acgtgggtcctctattaccagcttctaattgaattgtgttt 696

RESULT 13

AAF59611

ID AAF59611 standard; cDNA; 898 BP.

XX AAF59611;

AC AAF59611;

XX 24-APR-2001 (first entry)

DT Human cell cycle and proliferation protein CCYPR-22 cDNA, SEQ ID NO:76.

DE

XX Cell cycle and proliferation protein; CCYPR; human; agonist;

XX antagonist; gene therapy; detection; gene therapy;

KW transgenic animal disease model; immune disorder;

KW developmental disorder; cell signalling disorder;

KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;

KW arteriosclerosis; asthma; allergy; diabetes mellitus;

KW menstrual cycle disorder; bacterial infection; ss.

XX Homo sapiens.

OS

XX WO200107471-A2.

PN

XX 01-FEB-2001.

PD

XX 21-JUL-2000; 2000WO-US19948.

PF

XX 21-JUL-1999; 99US-0145075.

PR

XX 08-SEP-1999; 99US-0153129.

PR

XX 10-NOV-1999; 99US-0164647.

XX (INCY-) INCYTE GENOMICS INC.

PA

XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;

XX Azimzal Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;

PI WPI; 2001-112727/12.

DR

XX P-PSDB; AAB60474.

XX Human cell cycle and proliferation proteins and polynucleotides are

PT used to treat, diagnose and prevent immune, developmental and cell

PT signaling disorders and cell proliferative disorders including cancer -

XX

PS Claim 5; Page 181-182; 205pp; English.

XX

CC Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human

CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.

CC CCYPR and agonists of CCYPR are used to treat diseases or conditions

CC associated with decreased expression of functional CCYPR, while CCYPR

CC antagonists are used to treat diseases or conditions associated with

CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies

CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or

CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect

CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)

CC that specifically bind to CCYPR, and in drug screening methods to

CC identify compounds that modulate the activity of CCYPR. CCYPR

CC nucleotides can be used to generate transgenic animal models of human

CC disease, and can be used in gene therapy in target cells with genetic

CC abnormalities with respect to the expression of CCYPR for the

CC treatment or prevention of a disorder associated with CCYPR.

CC Diseases which can be diagnosed, treated and prevented using CCYPR

CC proteins, nucleic acids, agonists or antagonists include immune,

CC developmental and cell signalling disorders, and cell proliferative

CC disorders including cancer. Specific examples of these disorders

CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,

CC diabetes mellitus, disorders of the menstrual cycle and infections

CC caused by bacteria.

XX

Sequence 898 BP; 250 A; 186 C; 251 G; 211 T; 0 other;

Query Match 11.7%; Score 104; DB 22; Length 898;

Best Local Similarity 55.4%; Pred. No. 4.9e-21;

Matches 346; Conservative 0; Mismatches 240; Indels 38; Gaps 6;

Qy 283 aaaaacacacagaaaaaaatctcatctgcaaaatattcaccagagaaacgaagaga 342

Db 264 aagaggaaacgagctttaacaatctcatctgctggaataatgcaacaggaataatgatgaa 323

Qy 343 tgagcagcctatgcagaatggagaggaaga-----ccgcccttttggggaggag 390

Db 324 aagatgaaagagcaagttgctataaaggaggagcccttggccctacaccttgaaatgta 383

Qy 391 gtgaaggccacacgctgcaggaaga---tcgacggggacagagctgcgacttgcacctta 447

Db 384 gtgaatactgtgctcctagagaaacgtaggcggttcgcgcttagcagccatcctgc 443

Qy 448 atttcgattggccataccacatagagcagatcaatgatggatgggtgagatgagatg 507

Db 444 agtatagatgggacataatgcataggcttggagagccacaggaaggatgagagaggaga 503

Qy 508 atatgaaatattcattggaggagatgagagaaatcagaagaaacttagggagctgcagt 567

Db 504 atatgaaagattggggaggaggtgagacagctgatgaaagctgagggaaaaagcagt 563

Qy 568 tgaggaattgtctgcgtatcccttatggggagctctcttaatacaccatgaccatcatg 627

Db 564 tgagtcatagtctgcggcagtcagcactgatccccct---caccatgaccatcagctg 620

Qy 628 aatttgccttatgccttgactcctgccatttatcatgagataatactgtatctccgc 687

Db 621 agtttgccttatgccttgactcctgatggtttccctggaagttaatagggagaccctgc 680

Qy 688 tgtttctcttctcttgcatttctcctaataatgccttttactgacgttttgcgtggaacc 747

Db 681 ttccctaaacttaca-----catttgggtgacctttgtcgtaaacg 722

Qy 748 tatgtatttccatgtgtcaagtggtgttgcagagcttcttatttgaagattgcct 807

Db 723 ttttgatgttacctatttcttgggtctcctattaccagcttcttaaatgaaatgtgttt 782

Qy 808 ttg-cactcagtgtaagtttctgtcagcagtag-tttcaccctatttgcattggaataatt 865

Db 783 ttgacccagttgtgaagtttctgtcagcagggagagttttacattgtcatggaagatgc 842

Qy 866 aaagccataaagcaatttaaaaa 889

Db 843 tcattatattgtgaagttaata 866

RESULT 14

AAH75810

ID AAH75810 standard; cDNA; 767 BP.

XX AAH75810;

XX 17-OCT-2001 (first entry)

DT

XX





**THIS PAGE BLANK (USPTO)**

```
110 CAACCAAGAAAATCAAGAAAAGGAGCAAGTTGCTAATAAATAAGGGAGCCCT 159
34 euAlaLeuProLeuAspAlaGlyClnTyrCysValProArgGlyAsnArg 50
|||||
160 TGGCCCTCCCTTTGGATGCTGGTGAATACATGCTGCTAGAGGAATCGT 209
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
|||||
210 AGCGGTTCCCGCTTAGCGCCCATCTGCAGTATAGATGGGATATGAT 259
67 tHisArgLeuGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
260 GCATAGCTTTGGAGAACCAAGCAAGGATGAGAGAGAAATATGGAAA 309
84 rgileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
310 GGATTGGGAGAGGTGAGACACTGATGGAAGCTGAAGGAAAGCAG 359
101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117
|||||
360 TTGAGTCATAGTCTGGCGGAGTGCAGACTGACCCCTCACCATGACCA 409
117 sHisAspGluPheCysLeuMetPro 125
|||||
410 TCATGATGAGTTTGGCTTATGCCCC 434
seq_name: gb_est1:AI929703
seq_documentation_block: 576 bp mRNA EST 23-AUG-1999
LOCUS AI929703
DEFINITION au63f04.y1 Schneider fetal brain 00004: Homo sapiens cDNA clone
IMAGE: 2519455 5' similar to TR:088858 088858 REX-3. ;, mRNA
sequence.
ACCESSION AI929703
VERSION AI929703.1 GI:5665667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 576)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 444.
Location/Qualifiers
1..576
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2519455"
/clone.lib="Schneider fetal brain 00004"
/sex="male"
/tissue.type="frontal lobe"
/lab_stages="5 months post-conception"
/db_host="DHL08"
/Note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site 1: SstI; Site 2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
```

```
5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGCTGAGCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."
BASE COUNT 154 a 141 c 176 g 104 t 1 others
ORIGIN
alignment_scores:
Quality: 646.00 Length: 125
Ratio: 5.252 Gaps: 0
Percent Similarity: 98.400 Percent Identity: 97.600
alignment_block:
US-09-327-750D-32 x AI929703
Align seg 1/1 to: AI929703 from: 1 to: 576
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
202 ATGGAGTCCAAAGAGAAACGAGCAGTAAACAGTCTCAGCATGGAATGC 251
17 aAsnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
252 CAACCAAGAAAATCAAGAAAAGGAGCAAGTTGCTAATAAAGGGAGCCCT 301
34 euAlaLeuProLeuAspAlaGlyClnTyrCysValProArgGlyAsnArg 50
|||||
302 TGGCCCTCCCTTTGGATGCTGGTGAATACATGCTGCTAGAGGAATCGT 351
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
|||||
352 AGCGGTTCCCGCTTAGCGCCCATCTGCAGTATAGATGGGATATGAT 401
67 tHisArgLeuGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
402 GCATAGCTTTGGAGAACCAAGCAAGGATGAGAGAGAAATATGGAAA 451
84 rgileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
452 GGATTGGGAGAGGTGAGACACTGATGGAAGCTGAAGGAAAGCAG 501
101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117
|||||
502 TTGAGTCATAGTCTGGCGGAGTGCAGACTGACCCCTCACCATGACCA 551
117 sHisAspGluPheCysLeuMetPro 125
|||||
552 TCATGATGAGTTTGGCTTATGCCCC 576
seq_name: gb_est2:BG715659
seq_documentation_block: 734 bp mRNA EST 08-MAY-2001
LOCUS BG715659
DEFINITION 602676924F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:479976 5',
mRNA sequence.
ACCESSION BG715659
VERSION BG715659.1 GI:13994846
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 734)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
```

520 TCATGATGAGTTTTCCTTATGCCC 544

seq\_name: gb\_est1:AL526247

seq\_documentation\_block: 794 bp mRNA EST 13-FEB-2001  
 LOCUS AL526247 LTI\_NFL003\_NBC3 Homo sapiens cDNA clone CS0DC016Y001 5  
 DEFINITION prime, mRNA sequence.  
 ACCESSION AL526247  
 VERSION AL526247.1 GI:12789740  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 794)  
 AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

#### FEATURES

source

1..794  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DC016Y001"  
 /clone\_lib="LTI\_NFL003\_NBC3"  
 /sex="male"  
 /tissue\_type="neuroblastoma cells"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 236 a 163 c 213 g 180 t 2 others  
 ORIGIN

alignment\_scores:  
 Quality: 653.00 Length: 125  
 Ratio: 5.266 Gaps: 0  
 Percent Similarity: 99.200 Percent Identity: 98.400

#### alignment\_block:

US-09-327-750D-32 x AL526247

Align seg 1/1 to: AL526247 from: 1 to: 794

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetClnuAsnAl 17  
 148 ATGGAGTCCAAAGAGAAACGAGCAGTAAACAGTCTCAGCATGGAAATGC 197  
 17 aAsnGlnGluAsnGluGluLysGlnValAlaAsnLysGlyGluProL 34  
 198 CAACCAAGAAATGAAGAAAGGAGCAAGTGTCTATAAAGGGGAGCCCT 247  
 34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50  
 248 TGGCCCTCCCTTTGGATGCTGGTAACTACTGTGCTAGAGGAATCGT 297  
 51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTriaPleMe 67  
 298 AGCGGTTCCGCTTAGGAGGCCCATCCCTGCAGTATAGATGGATATGAT 347  
 67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84

348 GCATAGGCTTGGAGAACCCACAGGATGAGAGAAGACAATATGAAA 397

84 rgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100

398 GGATTGGGGAGGAGGTGAGACAGCTGATGGAAAGCTGAGGAAAGCAG 447

101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117

448 TTGAGTCATAGTCTGGGGCAGTCAGCACTGACCCCTCACCATGACCA 497

117 sHisAspGluPheCysLeuMetPro 125

498 TCATGATGAGTTTTCCTTATGCCC 522

seq\_name: gb\_est1:AV702643

seq\_documentation\_block: 683 bp mRNA EST 08-OCT-2000  
 LOCUS AV702643 ADB Homo sapiens cDNA clone ADBDBE11 5', mRNA sequence.  
 DEFINITION AV702643  
 ACCESSION AV702643  
 VERSION AV702643.1 GI:10718973  
 KEYWORDS EST.  
 SOURCE human.

#### ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

#### REFERENCE

AUTHORS

Peng, Y., Song, H., Huang, Q., Huang, C., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z. and Han, Z.

#### TITLE

Homo sapiens cDNA ADB clones

#### JOURNAL

Unpublished (2000)

#### COMMENT

Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@chgc.sh.cn  
 This clone is available at CHGC in Shanghai.

#### FEATURES

source

1..683  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="ADBDBE11"  
 /clone\_lib="ADB"  
 /tissue\_type="Adrenal gland"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 208 a 128 c 184 g 161 t 2 others  
 ORIGIN

alignment\_scores:  
 Quality: 650.00 Length: 125  
 Ratio: 5.242 Gaps: 0  
 Percent Similarity: 99.200 Percent Identity: 97.600

#### alignment\_block:

US-09-327-750D-32 x AV702643

Align seg 1/1 to: AV702643 from: 1 to: 683

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetClnuAsnAl 17  
 60 ATGGAGTCCAAAGAGAAACGAGCAGTAAACAGTCTCAGCATGGAAATGC 109  
 17 aAsnGlnGluAsnGluGluLysGlnValAlaAsnLysGlyGluProL 34



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4933012"
/clone_lib="NCI_CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with lp/19q loss"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      198 a   138 c   191 g   172 t
ORIGIN
alignment_scores:
  Quality: 653.00      Length: 125
  Ratio: 5.266        Gaps: 0
  Percent Similarity: 99.200      Percent Identity: 98.400
alignment_block:
US-09-327-750D-32 x BG820179
Align seg 1/1 to: BG820179 from: 1 to: 699
1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
94 ATGGAGTCCCAAGAGAAACGACGAGTAACAGTCTCAGCATGGAAATGC 143
17 aAsnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34
144 CAACCAAGAAATGAAGAAAGGAGCAAGTTGCTAATAAAGGGAGCCCT 193
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
194 TGGCCCTCCCTTTGGATGCTGGTGAATCTGTGCTATAGATGGATATGAT 243
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
244 AGCGCGTTCCGGCTTAGGAGCCCATCTGCAGTATAGATGGATATGAT 293
67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
294 GCATAGGCTTGGAGAACCAACAGCAAGGATGAGAGAAGAGAAATATGAAA 343
84 rGileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
344 GGATGGGAGGAGGTGACACAGCTGATGGAAGCTGAGGAAAGAGCAG 393
101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117
394 TTGAGTCATAGCTCGGGGACGTCAGCACTGACCCCTCCACCATGACCA 443
117 sHisAspGluPheCysLeuMetPro 125
444 TCATGATGAGTTTGGCTTATGCC 468
seq_name: gb_est1:BE790774

seq_documentation_block:
LOCUS      BE790774      783 bp      mRNA
DEFINITION 601581956F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3936613 5', mRNA sequence.
ACCESSION  BE790774
VERSION    BE790774.1 GI:10211972
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 783)
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)

```

## JOURNAL COMMENT

Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-re@mail.nih.gov  
 Tissue Procurement: DCFD/DRP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LICM780 row: 0 column: 14  
 High quality sequence start: 3  
 High quality sequence stop: 780.

## FEATURES

## source

```

1..783
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3936613"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
```

## BASE COUNT ORIGIN

```

214 a   169 c   219 g   181 t

```

## alignment\_scores:

```

  Quality: 653.00      Length: 125
  Ratio: 5.266        Gaps: 0
  Percent Similarity: 99.200      Percent Identity: 98.400

```

## alignment\_block:

```

US-09-327-750D-32 x BE790774

```

```

Align seg 1/1 to: BE790774 from: 1 to: 783

```

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
170 ATGGAGTCCCAAGAGAAACGACGAGTAACAGTCTCAGCATGGAAATGC 219
17 aAsnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34
220 CAACCAAGAAATGAAGAAAGGAGCAAGTTGCTAATAAAGGGAGCCCT 269
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
270 TGGCCCTCCCTTTGGATGCTGGTGAATCTGTGCTATAGATGGATATGAT 319
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
320 AGCGCGTTCCGGCTTAGGAGCCCATCTGCAGTATAGATGGATATGAT 369
67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
370 GCATAGGCTTGGAGAACCAACAGCAAGGATGAGAGAAGAGAAATATGAAA 419
84 rGileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
420 GGATGGGAGGAGGTGACACAGCTGATGGAAGCTGAGGAAAGAGCAG 469
101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117
470 TTGAGTCATAGCTCGGGGACGTCAGCACTGACCCCTCCACCATGACCA 519
117 sHisAspGluPheCysLeuMetPro 125

```

60 ATGGAGTCCAAAGAGAAACGACGAGTAAACAGTCTCAGCATGGAATAATGC 109  
 17 asnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGluProL 34  
 |||||  
 110 CAACCAAGAAATGAGAAAGAGCAAGTTGCTTAATAAGGGAGCCCT 159  
 |||||  
 34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50  
 |||||  
 160 TGGCCCTCCCTTTGGATGCTGCTGAATACTGTGCTAGAGGAATCGT 209  
 |||||  
 51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67  
 |||||  
 210 AGCGCGTTCCCGCTTAGGAGCCCATCTCTGCAGTATAGATGATGAT 259  
 |||||  
 67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGlu 84  
 |||||  
 260 GCATAGGCTTGGAGAACCCACAGCAGGATGAGAGAAGATAATGGAAA 309  
 |||||  
 84 rgileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100  
 |||||  
 310 GCATTGGGAGGAGGTGACACAGCTGTGGAAGAGCTGAGGCAAGACAG 359  
 |||||  
 101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117  
 |||||  
 360 TTGAGTCTAGTCTCGGGCAGCTCAGCACTGACCCCTCACCATTGACCA 409  
 |||||  
 117 shiAspGluPheCysLeuMetPro 125  
 |||||  
 410 TCATGATGAGTTTTCCTTATGCC 434

seq\_name: gb\_est1:AL526209

seq\_documentation\_block: 695 bp mRNA EST 13-FEB-2001  
 LOCUS AL526209 LTI\_NFL003\_NBC3 Homo sapiens cDNA clone CS0DC016Y001 3  
 DEFINITION AL526209 LTI\_NFL003\_NBC3 Homo sapiens cDNA clone CS0DC016Y001 3  
 prime, mRNA sequence.

ACCESSION AL526209  
 VERSION AL526209.1 GI:12789702  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 695)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)

CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
 Location/Qualifiers

1..695  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DC016Y001"  
 /clone\_lib="LTI\_NFL003\_NBC3"  
 /sex="male"  
 /tissue\_type="neuroblastoma cells"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library was normalized. Library was constructed  
 by Life Technologies. Contact : Feng Liang Life  
 Technologies, a division of Invitrogen 9800 Medical Center  
 Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
 8371 Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com"

BASE COUNT 158 a 198 c 154 g 183 t 2 others  
 ORIGIN

alignment\_scores:

Quality: 653.00 Length: 125  
 Ratio: 5.266 Caps: 0  
 Percent Similarity: 99.200 Percent Identity: 98.400

alignment\_block:

US-09-327-750D-32 x AL526209/rev ..

Align seg 1/1 to reverse of: AL526209 from: 1 to: 695

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
 |||||  
 546 ATGGAGTCCAAAGAGAAACGACGAGTAAACAGTCTCAGCATGGAATAATGC 497  
 |||||  
 17 asnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34  
 |||||  
 496 CAACCAAGAAATGAGAAAGAGCAAGTTGCTTAATAAGGGAGCCCT 447  
 |||||  
 34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50  
 |||||  
 446 TGGCCCTCCCTTTGGATGCTGCTGAATACTGTGCTAGAGGAATCGT 397  
 |||||  
 51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67  
 |||||  
 396 AGCGCGTTCCCGCTTAGGAGCCCATCTCTGCAGTATAGATGATGAT 347  
 |||||  
 67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGlu 84  
 |||||  
 346 GCATAGGCTTGGAGAACCCACAGCAGGATGAGAGAAGATAATGGAAA 297  
 |||||  
 84 rgileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100  
 |||||  
 296 GCATTGGGAGGAGGTGACACAGCTGTGGAAGAGCTGAGGCAAGACAG 247  
 |||||  
 101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117  
 |||||  
 246 TTGAGTCTAGTCTCGGGCAGCTCAGCACTGACCCCTCACCATTGACCA 197  
 |||||  
 117 shiAspGluPheCysLeuMetPro 125  
 |||||  
 196 TCATGATGAGTTTTCCTTATGCC 172

seq\_name: gb\_est2:BG820179

seq\_documentation\_block:

LOCUS BG820179 699 bp mRNA EST 22-MAY-2001  
 DEFINITION 602782222F1 NCI\_CGAP\_Brn67 Homo sapiens cDNA clone IMAGE:4933012  
 5', mRNA sequence.

ACCESSION BG820179  
 VERSION BG820179.1 GI:14167766  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 699)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Email: cgaps-r@mail.nih.gov  
 Tissue Procurement: David N. Louis, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM10859 row: 1 column: 05  
 High quality sequence stop: 695.

FEATURES  
 Location/Qualifiers  
 source 1..699

101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspH1 117  
 473 TTGAGTCATAGTCTGGCGCAGTCACACTGAGCCCTCACCATGACCA 522  
 117 sHisAspGluPheCysLeuMetPro 125  
 523 TCATGATGAGTTTGGCTTATGCCC 547

seq\_name: gb\_est1:AV702285

seq\_documentation\_block: 690 bp mRNA EST 08-OCT-2000  
 LOCUS AV702285 ADB Homo sapiens cDNA clone ADBBXE07 5', mRNA sequence.  
 DEFINITION AV702285 ADB Homo sapiens cDNA clone ADBBXE07 5', mRNA sequence.  
 ACCESSION AV702285  
 VERSION AV702285.1 GI:10718615  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Peng, Y., Song, H., Huang, Q., Gu, Y., Yang, Y., Gao, G., Xiao  
 , H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z.,  
 Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu  
 , G., Hu, R., Chen, J., Chen, Z. and Han, Z.  
 TITLE Homo sapiens cDNA ADB clones  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@chgc.sh.cn  
 This clone is available at CHGC in Shanghai.

FEATURES  
 source  
 1. .690  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="ADBBXE07"  
 /clone\_lib="ADB"  
 /tissue\_type="Adrenal gland"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
 XhoI"  
 BASE COUNT 210 a 132 c 180 g 166 t 2 others  
 ORIGIN

alignment\_scores:  
 Quality: 653.00 Length: 125  
 Ratio: 5.266 Gaps: 0  
 Percent Similarity: 99.200 Percent Identity: 98.400  
 alignment\_block:  
 US-09-327-750D-32 x AV702285 ..  
 Align seg 1/1 to: AV702285 from: 1 to: 690

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
 60 ATGGAGTCCAAAGAGAAACGAGCAGTAAACAGTCTCAGCATGGAATGC 109  
 17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34  
 110 CAACCAAGAAATGAAGAAAGGAGCAGTTCCTAATAAGGGAGGCCCT 159  
 34 euAlaLeuProLeuAspAlaGlyGlyTyrCysValProArgGlyAsnArg 50  
 160 TGCCCTCCCTTTGGATGCTGGGTAATACTGTGTGCTAGAGGAATCGT 209  
 51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67

210 AGGCGGTTCCGGCTTAGGCAGGCCCATCTCCAGTATAGATGGATGAT 259  
 67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84  
 260 GCATAGGCTTGGAGAACACAGCAGGATGAGAGAGAGATATATGAAA 309  
 84 rgileGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100  
 310 GGATTGGGAGGAGGTGAGACAGCTGATGAAAAGCTGAGGGAAGACAG 359  
 101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspH1 117  
 360 TTGAGTCATAGTCTGGCGCAGTCACACTGAGCCCTCACCATGACCA 409  
 117 sHisAspGluPheCysLeuMetPro 125  
 410 TCATGATGAGTTTGGCTTATGCCC 434  
 seq\_name: gb\_est1:AV705808  
 seq\_documentation\_block: 691 bp mRNA EST 09-OCT-2000  
 LOCUS AV705808 ADB Homo sapiens cDNA clone ADBCRH03 5', mRNA sequence.  
 DEFINITION AV705808 ADB Homo sapiens cDNA clone ADBCRH03 5', mRNA sequence.  
 ACCESSION AV705808  
 VERSION AV705808.1 GI:10723103  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Peng, Y., Song, H., Huang, Q., Gu, Y., Yang, Y., Gao, G., Xiao  
 , H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z.,  
 Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu  
 , G., Hu, R., Chen, J., Chen, Z. and Han, Z.  
 TITLE Homo sapiens cDNA ADB clones  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@chgc.sh.cn  
 This clone is available at CHGC in Shanghai.  
 FEATURES  
 source  
 1. .691  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="ADBCRH03"  
 /clone\_lib="ADB"  
 /tissue\_type="Adrenal gland"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
 XhoI"  
 BASE COUNT 211 a 133 c 181 g 162 t 4 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 653.00 Length: 125  
 Ratio: 5.266 Gaps: 0  
 Percent Similarity: 99.200 Percent Identity: 98.400  
 alignment\_block:  
 US-09-327-750D-32 x AV705808 ..  
 Align seg 1/1 to: AV705808 from: 1 to: 691  
 1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
 60 ATGGAGTCCAAAGAGAAACGAGCAGTAAACAGTCTCAGCATGGAATGC 109  
 17 aAsnGlnGluAsnGluLysGluGlnValAlaAsnLysGlyGluProL 34  
 110 CAACCAAGAAATGAAGAAAGGAGCAGTTCCTAATAAGGGAGGCCCT 159  
 34 euAlaLeuProLeuAspAlaGlyGlyTyrCysValProArgGlyAsnArg 50  
 160 TGCCCTCCCTTTGGATGCTGGGTAATACTGTGTGCTAGAGGAATCGT 209  
 51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67

http://image.llnl.gov  
 Plate: LIAM10003 row: d column: 14  
 High quality sequence stop: 658.  
 Location/Qualifiers  
 source 1..658

## FEATURES

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4361053"  
 /clone\_lib="NIH\_MGC\_84"  
 /tissue\_type="adrenal cortex carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

BASE COUNT 205 a 124 c 164 g 165 t  
 ORIGIN

alignment\_scores:  
 Quality: 653.00 Length: 125  
 Ratio: 5.266 Gaps: 0  
 Percent Similarity: 99.200 Percent Identity: 98.400

## alignment\_block:

US-09-327-750D-32 x BF971303 ..

Align seg 1/1 to: BF971303 from: 1 to: 658

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17  
 |||||

24 ATGGAGTCCAAAGAGAAACGACGAGTAACAGCTCTCAGCATGGAAATGC 73

17 aasnGlnGluAsnGluGluGluGluValAlaAsnLysGlyGluProL 34

74 CAACCAAGAAATGAAGAAAGGACGACAGTTGCTAATAAAGGGAGCCCT 123

34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50

124 TGGCCCTCCCTTTGGATGCTGGTGAATCTGTGCTAGAGGAATCGT 173

51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67

174 AGCGGTTCCCGGTTAGGCGAGCCCATCTGCAGTATAGATGGATATGAT 223

67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84

224 GCATAGGCTTGGAGAACCCACAGGCAAGGATGAGAGAAGAGATATGGAA 273

84 rgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100

274 GGATGGGAGAGGTGAGACAGCTGTGGAAGAGCTGAGGGAAGAGCAG 323

101 LeuSerHisSerLeuArgAlaValSerThrAspProProHisHisAspHi 117

324 TTGAGTCATAGTCTCGGGCAGTCAGCACTGACCCCTCACCATGACCA 373

117 shiAspGluPheCysLeuMetPro 125

374 TCATGATGAGTTTGCCTTATGCC 398

seq\_name: gb\_est2:BG714823

seq\_documentation\_block:

LOCUS BG714823 674 bp mRNA 08-MAY-2001  
 DEFINITION 602677146F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4799747 5',  
 mRNA sequence.

ACCESSION BG714823

VERSION BG714823.1 GI:13993754

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 674)  
 NIH-MGC http://mgs.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: LIAM10689 row: k column: 12  
 High quality sequence stop: 672.

## FEATURES

source

1..674  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4799747"  
 /clone\_lib="NIH\_MGC\_96"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pBluescriptR (modified  
 pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.3 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 175 a 156 c 201 g 142 t  
 ORIGIN

## alignment\_scores:

Quality: 653.00 Length: 125  
 Ratio: 5.266 Gaps: 0  
 Percent Similarity: 99.200 Percent Identity: 98.400

## alignment\_block:

US-09-327-750D-32 x BG714823 ..

Align seg 1/1 to: BG714823 from: 1 to: 674

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17

173 ATGGAGTCCAAAGAGAAACGACGAGTAACAGCTCTCAGCATGGAAATGC 222

17 aasnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34

223 CAACCAAGAAATGAAGAAAGGACGACAGTTGCTAATAAAGGGAGCCCT 272

34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50

273 TGGCCCTCCCTTTGGATGCTGGTGAATCTGTGCTAGAGGAATCGT 322

51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67

323 AGCGGTTCCCGGTTAGGCGAGCCCATCTGCAGTATAGATGGATATGAT 372

67 thisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84

373 GCATAGGCTTGGAGAACCCACAGGCAAGGATGAGAGAAGAGATATGGAA 422

84 rgIleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100

423 GGATGGGAGAGGTGAGACAGCTGATGGAAAGCTGAGGAAAGAGCAG 472

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
143 ATGGAGTCCAAAGAGAAACGAGCAGTAAACAGTCTCAGCATGGAAATGC 192
|||||
17 aasnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
193 CAACCAAGAAATGAAGAAAGGAGCAAGTTGCTAATAAAGGGAGCCCT 242
|||||
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
243 TGGCCCTCCCTTTGGATGCTGGTGAATCTGTGCTAGAGGAAATCGT 292
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
|||||
293 AGCGCGTTCCGCTTAGGCAGCCCATCTGCAGTATAGATGGATATGAT 342
|||||
67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
343 GCATAGGCTTGGAGAACACAGCAGCAAGGATGAGAGAGAGAAATATGAAA 392
|||||
84 rGleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
393 GGATGGGAGGAGGTGAGACAGCTGATGAAAGCTGAGGAGAAAGCAG 442
|||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117
|||||
443 TTGAGTCATAGTCTCGGCGAGTCAGCAGCTGACCCCTCACCATGACCA 492
|||||
117 sHisAspGluPheCysLeuMetPro 125
|||||
493 TCATGATGAGTTTTCCTTATGCCC 517
|||||

```

seq\_name: gb\_est2:BG472167

```

seq_documentation_block: 634 bp mRNA EST 21-MAR-2001
LOCUS BG472167
DEFINITION 60251389AF1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4645775 5',
mRNA sequence.
ACCESSION BG472167
VERSION BG472167.1 GI:13404541
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM419 row: k column: 24
High quality sequence stop: 634.
Location/Qualifiers
1..634
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4645775"
/clone_lib="NIH_MGC_16"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pOTB7; Site:1; XhoI: Site:2;
EcoRI: cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGCAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit

```

FEATURES  
source

(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library.

BASE COUNT 187 a 119 c 166 g 162 t  
ORIGIN

alignment\_scores:  
Quality: 653.00 Length: 125  
Ratio: 5.266 Gaps: 0  
Percent Similarity: 99.200 Percent Identity: 98.400

alignment\_block:

US-09-327-750D-32 x BG472167 ..

Align seg 1/1 to: BG472167 from: 1 to: 634

```

1 MetGluSerLysGluLysArgAlaValAsnSerLeuSerMetGluAsnAl 17
|||||
30 ATGGAGTCCAAAGAGAAACGAGCAGTAAACAGTCTCAGCATGGAAATGC 79
|||||
17 aasnGlnGluAsnGluGluLysGluGlnValAlaAsnLysGlyGluProL 34
|||||
80 CAACCAAGAAATGAAGAAAGGAGCAAGTTGCTAATAAAGGGAGCCCT 129
|||||
34 euAlaLeuProLeuAspAlaGlyGluTyrCysValProArgGlyAsnArg 50
|||||
130 TGGCCCTCCCTTTGGATGCTGGTGAATCTGTGCTAGAGGAAATCGT 179
|||||
51 ArgArgPheProValArgGlnProIleLeuGlnTyrArgTrpAspIleMe 67
|||||
180 AGCGCGTTCCGCTTAGGCAGCCCATCTGCAGTATAGATGGATATGAT 229
|||||
67 tHisArgLeuGlyGluProGlnAlaArgMetArgGluGluAsnMetGluA 84
|||||
230 GCATAGGCTTGGAGAACACAGCAGCAAGGATGAGAGAGAGAAATATGAAA 279
|||||
84 rGleGlyGluGluValArgGlnLeuMetGluLysLeuArgGluLysGln 100
|||||
280 GGATGGGAGGAGGTGAGACAGCTGATGAAAGCTGAGGAGAAAGCAG 329
|||||
101 LeuSerHisSerLeuArgAlaValSerThrAspProHisHisAspHi 117
|||||
330 TTGAGTCATAGTCTCGGCGAGTCAGCAGCTGACCCCTCACCATGACCA 379
|||||
117 sHisAspGluPheCysLeuMetPro 125
|||||
380 TCATGATGAGTTTTCCTTATGCCC 404
|||||

```

seq\_name: gb\_est2:BF971303

```

seq_documentation_block: 658 bp mRNA EST 22-JAN-2001
LOCUS BF971303
DEFINITION 602273150F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4361053 5',
mRNA sequence.
ACCESSION BF971303
VERSION BF971303.1 GI:12338518
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 658)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

